DOCTOR OF PHILOSOPHY

Structure and molecular recognition in riboswitches

Daldrop, Peter

Award date:
2011

DOCTOR OF PHILOSOPHY

Structure and molecular recognition in riboswitches

Peter Daldrop

2011

University of Dundee

University of Dundee

College of life Sciences

Structure and Molecular Recognition in Riboswitches

Peter Daldrop

A thesis submitted for the degree of:

Doctor of Philosophy
University of Dundee

December 9, 2011

Declarations

I declare that the following thesis is based on the results of investigations conducted by myself, and that this thesis is of my own composition. Work other than my own is clearly indicated in the text by reference to the relevant researchers or to their publications. This dissertation has not in whole, or in part, been previously submitted for a higher degree.

Peter Daldrop

I certify that Peter Daldrop has spent the equivalent of at least nine terms in research work at the College of Life Sciences, University of Dundee, and that he has fulfilled the conditions of the Ordinance General No. 14 of the Universtiy of Dundee and is qualified to submit the accompanying thesis in application for the degree of Doctor of Philosophy.

Dr. Ruth Brenk

Acknowledgments

I would like to thank Dr. Ruth Brenk for the opportunity to work in her group on this project which I enjoyed very much. Her scientific expertise and guidance has been a great help in these years and in preparation of this dissertation.

Also I would like to thank Dr. David Lilley, who acted as my co-supervisor. I am grateful for the opportunity to benefit from his knowledge and experience in RNA research, the wonderful collaboration especially for the project which led to the k-turn chapter. His comments on the manuscript were of great help. Additionally, I would like to thank him for offering me a postdoctoral position in his lab.

Special thanks deserve Dr. Kersten Schroeder and Colin Hammond who contributed results to this dissertation and with whom I enjoyed collaborating very much.

Dr. David Robinson deserves thanks for his patience in answering all my questions regarding crystallography and who was of great help in data analysis.

Thanks also to the people that read parts of the thesis ,helped to spot mistakes and with their suggestions helped a lot in improving the first draft to this final version.

Additionally, I would like to thank all members of the Brenk and Lilley labs, past and present, for their support, help and all the coffee breaks.

I also would like to thank Dr. Robert T. Batey for his generous gift of plasmids and his support in crystallizing the riboswitches. I really enjoyed visiting his lab in 2008 and would also like to thank him and his whole group for the hospitality I experienced.

I would like to thank my family for emotional support and distraction when needed.

Special thanks goes to my girlfriend, Silke, for her support and love, as well as for tolerating my absence from Berlin for the past 4 years.

Thanks to my friends in Dundee and elsewhere for their support, friendship and patience. In particular thanks to Helena Salmen my friend and flatmate for all the time spent together, especially walking up Scottish munros in varying weather!

Finally I would like to thank the Wellcome Trust for providing the funding.

Abstract

Riboswitches are *cis*-acting gene regulatory RNAs, which function without involvement of proteins. They have been implicated as drug targets and are attractive systems for the study of RNA-ligand binding and RNA folding.

The purine riboswitch was used as a model system for RNA-ligand docking. Published binding data was successfully reproduced *in silico* and compounds predicted to bind the riboswitch in a virtual screening were tested experimentally. Structural data confirming the predicted binding mode for several cases was obtained. The problems encountered were not specific to RNA-ligand docking but known from the far more explored field of protein-ligand docking.

The SAM-I riboswitch was also subjected to virtual ligand screening. This receptor is a system of greater complexity than the purine riboswitch and consequently posed a harder challenge to the docking protocol. After initial validation of the docking setup based on previously published data, a set of compounds selected from the in-house database of commercially available compounds was screened. One compound identified *in silico* was confirmed to bind experimentally.

The k-turn motif found in the SAM-I riboswitch was investigated with respect to its folding. The k-turn motif was found to be foldable in context of the SAM-I riboswitch as well as in isolation as was expected. Furthermore, mutations disrupting key interactions within the k-turn motif were found to be prohibitive of k-turn folding in isolation as well as in context of the riboswitch, leading to a loss of ligand binding. Interestingly, two sequences were identified which fold in context of the riboswitch but do not fold in isolation. This confirms the contribution of tertiary interactions to k-turn folding. This conclusion was backed up with structural data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Antibiotics and the need for new drugs

1.1.1 Antibiotics

The term antibiotic was coined in the late 40s of the last century by Selman A. Waksman (Waksman, 1947): "An antibiotic is a chemical substance, produced by micro-organisms, which has the capacity to inhibit the growth of or even to destroy bacteria and other micro-organisms."

Nowadays antibiotics can include synthetic or semi synthetic substances, thus not requiring to come from a microbial source. Antibiotics have been known since the 1920s, and have had a huge impact on human health (Demain, 2009). Before antibiotics were available, infectious diseases were the most common cause of death. Millions of patients succumbed to infections like pneumonia and gastroenteritis, which today are usually easily treatable by antibiotics.

One of the first major antibiotics, penicillin, is a substance isolated from *Penicillium* fungi (Fleming, 1929) and it is still used to treat bacterial infections in humans. Later, synthetic penicillin and slightly modified molecules were introduced as well as other naturally occurring substances and derivatives of those (Demain, 2009). Most antibiotics used today are derivatives of naturally oc-

curring substances and act on a number of targets essential in bacterial growth. Penicillin is an inhibitor of cell wall synthesis, but other antibiotic targets include DNA gyrase (Pommier et al., 2010), DNA dependent RNA polymerase (McClure and Cech, 1978) and the ribosome (Hermann, 2005).

1.1.2 The need for new drugs

The effort to discover new drugs, in particular antibiotics, needs to continue for two reasons:

1. There are still diseases for which there is no satisfactory treatment available
2. Resistance against many commonly used antibiotics has emerged and thus they become ineffective as a consequence.

Due to the extensive use of antimicrobial drugs, resistant strains of pathogens have emerged and pose a serious health threat (Hawkey, 2008). In hospital environments, a large fraction of the occurring infections are caused by bacteria such as *Staphylococcus aureus*, that are resistant to several or virtually all available antibiotics for treatment (Hawkey, 2008). For a long time vancomycin served as the last resort in these cases, since no resistant strains were known and this drug was used as a means of last resort. However, in 1987 vancomycin resistant strains of bacteria were found (Lütticken and Kunstmann, 1988) and more cases of vancomycin resistant pathogens have emerged since. For a number of pathogens, e.g. *Mycobacterium tuberculosis*, and *S. aureus*, strains have been isolated from patients that are resistant to all antibiotics used for treatment. Therefore, the efforts in drug discovery for infectious diseases must continue to ensure effective treatment options in the future. One possible path is to explore new targets and make these available for structure-based drug discovery. These new targets could be found in functional RNA entities.

1.2 RNA

1.2.1 RNA in biology and disease

The field of RNA research has advanced hugely over the last years. RNA molecules have been shown to be essential factors in almost every process in living cells. The new functions discovered for RNA molecules have changed the perception of RNA. Instead of being viewed as a class of static messenger molecules, RNA is now widely regarded as an essential and active contributor in processes like transcriptional (Bayne and Allshire, 2005; Winkler et al., 2003) and translational (Jovanovic and Hengartner, 2006; Sen and Blau, 2006) regulation, protein function (Manche et al., 1992; Sledz et al., 2003) and catalysis (Fedor and Williamson, 2005; Lilley, 2011; Thomas and Hergenrother, 2008).

Additionally, the importance of RNA factors in many disease states has been demonstrated. The portfolio of diseases with significant involvement of RNA factors includes cancer (Muniyappa et al., 2009), diabetes mellitus (Guay et al., 2011) and infectious diseases (Huang et al., 2011). A special class of RNA-related diseases is represented by the RNA viruses which pose a serious threat to human health. RNA viruses store their genomic information in RNA form rather than in DNA form. Among the most prominent of these viruses are the human immunodeficiency virus (HIV; Frankel and Young 1998), the hepatitis C virus (HCV; Ashfaq et al. 2011) and the influenza viruses.

1.2.2 RNA as a drug target

RNA molecules offer a valid target for intervention into disease related biological processes. Many functions of RNA in biologically-important processes do not involve proteins and even in RNA-protein complexes, the RNA component is often essential (Fedor and Williamson, 2005; Autexier and Lue, 2006).

In the last decades all RNA-targeting drugs used in the clinic have been antibi-

otics targeting the large or the small subunit of the ribosome (Lin et al., 1997; Bozdogan and Appelbaum, 2004; Hermann, 2005).

In the ribosome, the RNA constitutes the active component for both proof-reading (Yoshizawa et al., 1999) and peptide bond formation (Rodnina et al., 2007). Consequently, it is unsurprising that many antibiotics targeting the ribosome do so by acting on the ribosomal RNA of both the small and the large ribosomal subunit (Hermann, 2005). For example, tetracycline binds to the small ribosomal subunit preventing the binding of aminoacyl-tRNA (Brodersen et al., 2000). Macrolide antibiotics have been shown to bind the large ribosomal subunit, more specifically the ribosomal RNA (Hansen et al., 2002).

The antimicrobials targeting RNA prove, that manipulation of cellular processes via RNA targets is possible. Most of the antimicrobials of this category are not molecules with "drug-like" properties (Lipinski, 2001). Aminoglycosides and macrolide antibiotics violate the "rule of five" in several ways. They exceed a molecular weight of 500, contain more than 5 hydrogen bond donors and 10 hydrogen bond donors. Additionally, these antibiotics contain a large amount of stereo centers, which in conjunction with their size makes them difficult to synthesize. However, it has been shown that RNA molecules can bind virtually any type of ligand, implying that drug-like molecules can be used in principle to interfere with RNA function. RNAs specifically binding very diverse chemicals have been generated by *in vitro* selection (Tuerk and Gold, 1990). Among these ligands for which RNA aptamers have been generated (Fig. 1.1) are theophylline (Jenison et al., 1994), citrulline, valine, tryptophan and flavin mononucleotide (FMN; Wilson and Szostak 1999).

In the light of these data it is unsurprising that in addition to the ribosome several other functional RNA species have been proposed as drug targets. Some prominent examples are tRNAs (Walter et al., 2002), the HIV-1 TAR element (human immunodeficiency virus 1 trans-activation response element; Frankel and Young 1998; Lind et al. 2002), riboswitches (Blount et al., 2007; Lee et al., 2009;

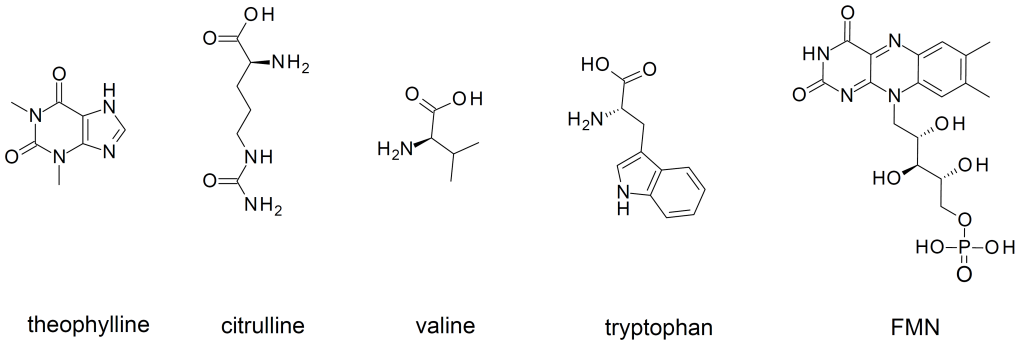


Figure 1.1: Ligands for which artificial aptamers have been isolated. These ligands include nucleobase analogs like theophylline as well as amino acids and FMN.

Mulhbacher et al., 2010) and ribozymes (Mulhbacher et al., 2010).

1.2.3 Computational drug discovery and RNA

Drug discovery is the process of identifying, validating and testing a chemical compound as a therapeutic agent against a selected medical condition. The identification of candidate chemicals binding a molecular target is one of the first hurdles that must be overcome. In modern drug discovery *in silico* methods such as virtual ligand screening have become a very important tool for the discovery of novel ligands for any given target (Ripphausen et al., 2010). Among the various virtual screening techniques, molecular docking has proven to be a powerful method for the discovery of novel ligands from a large set of candidate compounds. Molecular docking predicts the three-dimensional binding mode of given compounds to the receptor binding site and estimates the binding affinity. A plethora of docking algorithms and scoring functions have been developed in the past and excellent reviews on the topic are available (Klebe, 2006; McInnes, 2007; Yuriev et al., 2011). While docking has been extensively applied to proteins, only few docking studies for RNA targets have been published and studies including experimental validation of obtained docking results are even rarer (Fulle and Gohlke, 2010). However, with the available structural data that emerged in

recent years, structure based computational methods are being applied to RNA targets (Franceschi and Duffy, 2006; Schwalbe et al., 2007; Serganov, 2010). The studies of RNA-ligand docking recently reported follow one of two approaches:

1. adaptation of methods and scoring functions originally developed for protein ligand docking (Lind et al., 2002; Detering and Varani, 2004; Kang et al., 2004; Moitessier et al., 2006; Park et al., 2008; Lang et al., 2009; Li et al., 2010)

2. development of entirely new methods tailored to RNA targets (Morley and Afshar, 2004; Barbault et al., 2006; Pfeffer and Gohlke, 2007; Guilbert and James, 2008)

Most of these studies attempt to reproduce available binding data, with varying degrees of success. For the newly developed methods reproduction of previously described binding modes and enrichment of known ligands from a large set of compounds with unknown binding properties have been reported. Morley and Afshar 2004 use RiboDock with an empirical scoring function. This scoring function is then validated by reproduction of binding modes of reported NMR solution structures and enrichment factor experiments for the TAR RNA. Barbault et al. 2006 use a neural neural network to develop a scoring function to use with the Autodock software. The developed method is then used to reproduce binding data for RNA-aminoglycoside complexes. Pfeffer and Gohlke 2007 follow a slightly different approach by deriving a knowledge based scoring function based on several hundred nucleic acid - small molecule and nucleic acid - protein complex crystal structures. The resulting scoring function is then validated by prediction of binding affinities for 15 RNA-ligand complexes. Guilbert and James 2008 developed a software called 'MORDOR' and validated it against a set of 57 RNA-ligand complex structures from the PDB. For 74 % of this set the binding mode could be predicted with an RMSD of $< 2.5\ \mathring{A}$. Using the 'MORDOR' software (Pinto et al., 2008) reported the discovery of novel ligands for

human telomerase RNA with affinities between 180 μM and 1.5 mM. However no structural confirmation of the reported binding modes was obtained. For the application of docking software from the protein field the situation is similar. Most well-known docking programs from the protein field have been applied to RNA targets: DOCK (Detering and Varani, 2004; Kang et al., 2004; Moitessier et al., 2006), DOCK6 (Lang et al., 2009) AutoDock (Detering and Varani, 2004), FlexX (Park et al., 2008), Glide (Li et al., 2010) and Gold (Li et al., 2010). These studies with the exception of Park et al. 2008 only report predictions for ligands already known to bind. Park et al. 2008 additionally report novel compounds for the selected target, a pseudoknot inducing ribosomal frame shifting. The binding of these compounds is determined by a frame shifting assay and no K_D values are reported. Many of the used complexes bind large and flexible ligands, water-mediated interactions and extremely flexible receptors. These characteristics still pose a considerable challenge in the far more explored protein-ligand docking field (Klebe, 2006) and thus it is difficult to disentangle effects specific to RNA-ligand docking from problems generic to all known docking methods. Most importantly, no study which includes validation against known data, identification of novel ligands and confirmation of the predicted binding mode by structural methods. To drive molecular docking for RNA targets forward, an amenable system is required which allows thorough method testing. If such a system offers potential applications for the discovered ligands, this would be an additional incentive for investigating this target.

Riboswitches are a class of naturally occurring RNA molecules which have the potential to serve this purpose. They have been proposed as drug targets (Blount and Breaker, 2006; Kim et al., 2009; Lee et al., 2009; Ott et al., 2009; Mulhbacher et al., 2010; Deigan and Ferre-D'Amare, 2011), since they control important metabolic pathways in bacteria depending on ligand binding, while not occurring in humans. Therefore in addition to their use as a model system they may be of therapeutic importance.

1.3 Riboswitches

1.3.1 General principle

Riboswitches are *cis*-acting regulatory RNA elements involved in gene regulation. The typical riboswitch is part of the 5' leader sequence of mRNA. Located upstream of the start codon of the regulated gene, the riboswitch acts as a molecular switch for the expression of the genes encoded in the particular mRNA. Typically, the riboswitch is divided into two distinct functional domains, the aptamer and the expression platform (Tucker and Breaker, 2005). The aptamer domain acts as the sensory domain, and adopts different conformations in the presence absence of the specific ligand (Nudler and Mironov, 2004). This information is then relayed to the expression platform, which converts the signal into an alteration of gene expression (Nudler and Mironov, 2004; Rieder et al., 2007; Wickiser et al., 2005; Barrick and Breaker, 2007).

The ligand recognized by the aptamer domain is specific for each riboswitch and represents a metabolite like adenine (Mandal and Breaker, 2004), guanine (Batey et al., 2004), S-adenosylmethionine (SAM; Winkler et al. 2003), cobalamine (Nahvi et al., 2004) or lysine (Sudarsan et al. 2003b; Fig. 1.2). Riboswitches sensing the second messenger cyclic-di-GMP (Fig. 1.2) have also been reported (Sudarsan et al., 2008).

The aptamer domain can bind the ligand tightly; the *in vivo* ligand of many riboswitches has an affinity below $K_D = 1\ \mu M$ in *in vitro* binding experiments (Montange et al., 2010; Gilbert et al., 2006a; Winkler et al., 2002). Ligand binding stabilizes a well defined three-dimensional conformation of the aptamer domain. For many riboswitches the aptamer domain in complex with the ligand has been crystallized and the three-dimensional structure has been determined (Edwards and Ferré-D'Amaré, 2006; Montange and Batey, 2006; Batey et al., 2004; Garst et al., 2008; Smith et al., 2009, 2011). In the absence of the ligand the structure of the aptamer domain is usually less defined, and the apo-structure of several

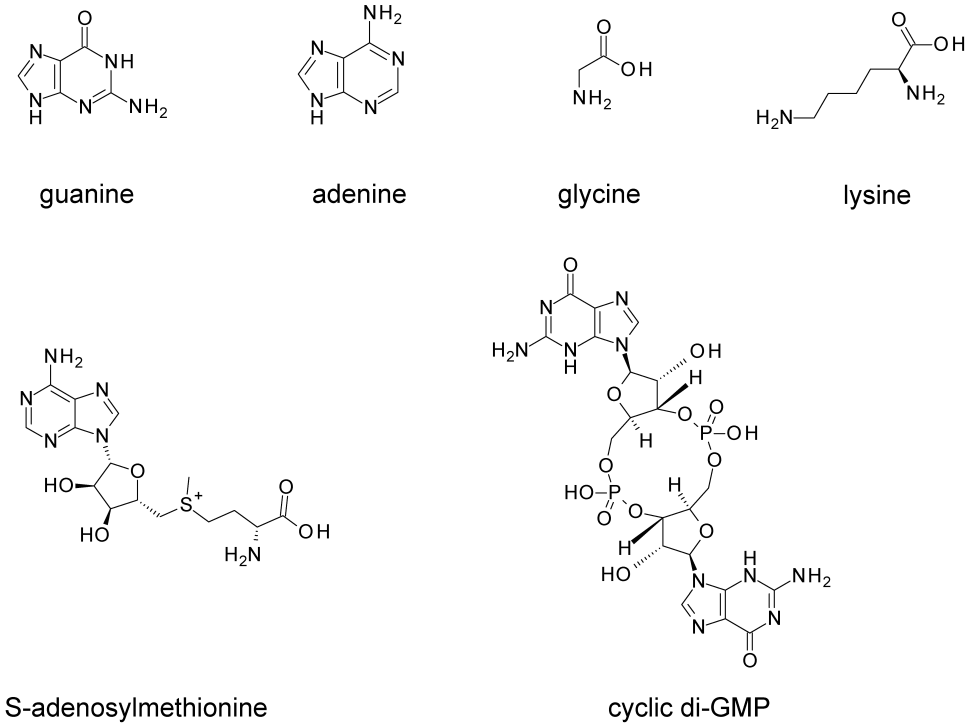


Figure 1.2: Chemical compounds that are naturally occurring riboswitch ligands. These include metabolites such as nucleobases (adenine, guanine) as well as amino acids (lysine, glycine) and the central metabolite SAM. Also the second messenger cyclic di-GMP was found to be a natural riboswitch ligand.

aptamer domains have been determined by X-ray crystallography. Among these are the SAM-I riboswitch (Stoddard et al., 2010), the lysine riboswitch (Garst et al., 2008) and the FMN riboswitch (Vicens et al., 2011).

The expression platform needs to translate the conformational signal from the aptamer domain into an alteration of gene expression. Two principal classes of logical switches are conceivable: the ON-switch and the OFF-switch (Fig. 1.3). The OFF-switch prevents gene expression in the ligand bound state while permitting gene-expression in the ligand-free state (Fig. 1.3 A). The ON-switch (Fig. 1.3 B) works the opposite way. For both classes examples can be found occurring in bacteria. Interestingly, the adenine riboswitch of *Bacillus subtilis* is an example for an ON-switch (Lemay et al., 2011) while the guanine riboswitch of *B. subtilis* acts as an OFF-switch. The aptamer domains of both riboswitches are

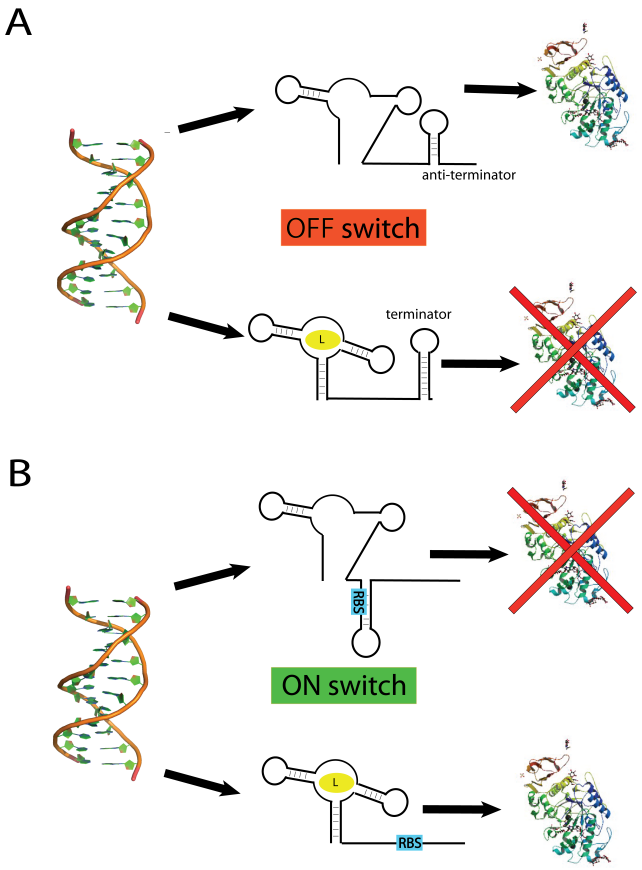


Figure 1.3: Logical types of riboswitches. A: OFF-switch by transcription termination; DNA is transcribed to RNA which in the absence of ligand results in formation of an anti-terminator stem loop and subsequent protein expression. In the case of ligand binding transcription is prevented by a terminator stem loop and protein production is prevented. B: ON-switch; DNA is transcribed to RNA and in the case of no ligand binding to the riboswitch the ribosomal binding site (RBS) is accessible resulting in protein production. In the absence of ligand ribosomal binding to the mRNA is prevented by sequestering of the RBS into a stem loop, preventing protein production.

highly similar and their binding sites are virtually identical except for one single base. This nucleobase, a uridine in the case of the adenine riboswitch, forms Watson-Crick type hydrogen bonds with the ligand adenine. In the guanine riboswitch this uridine residue is replaced by a cytidine, also forming Watson-Crick type hydrogen bonds with the ligand guanine. Whether a riboswitch is part of the ON- or OFF-switch category is determined by the expression platform.

1.3.2 Mechanisms of riboswitches

There are two different aspects of gene regulation by riboswitches that need to be discussed in a mechanistic context; first the mechanism of protein expression control and second the mechanism of ligand sensing. There are four known fundamental mechanisms by which riboswitch control can occur:

1. transcription attenuation
2. translation attenuation
3. anti-sense mechanism
4. ribozyme cleavage

A riboswitch employing the transcription attenuation mechanism regulates gene expression by preventing transcription of the full mRNA in the OFF state of the switch. This usually happens by formation of a GC-rich terminator stem (Fig. 1.3 A; Mandal and Breaker 2004; Rodionov et al. 2003), which is followed by a poly-U stretch (Barrick and Breaker, 2007). This stem causes the RNA-polymerase to dissociate and results in an incomplete transcript RNA not coding for a functional protein (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999). The second mechanism is translation attenuation, mediated via structural changes in the mRNA that block translation initiation. In the simplest case this is achieved by sequestering the ribosomal binding site into an element of secondary structure abolishing ribosome binding (Fig. 1.3 B; Rodionov et al. 2002; Vitreschak et al. 2004). Regulation by an antisense transcript is a variation of the transcription attenuation mechanism. The transcription of the antisense RNA is controlled by the riboswitch, and the full transcript then inhibits the translation of the target gene. An example for such a *trans*-riboswitch is the SreA riboswitch of *Listeria monocytogenesis* (Loh et al., 2009). Antisense RNAs in general as regulatory elements in prokaryotes have been reported and described (Gottesman, 2002).

Regulation by ribozyme cleavage has up to this date only been found in one case, the *glmS* riboswitch/ribozyme. The aptamer domain binds glucosamine-6-phosphate and utilizes it as a cofactor for self-cleavage (Winkler et al., 2004). The such cleaved mRNA is subsequently targeted for degradation, preventing gene expression (Collins et al., 2007).

The first two mechanisms are widely used in riboswitch gene regulation while the third and fourth are only known to occur in a single example each (Rodionov et al., 2004; Winkler et al., 2004).

For the mechanism of ligand sensing two scenarios have been observed, where riboswitches are under thermodynamic or kinetic control. For a riboswitch under thermodynamic control the fraction of riboswitch in the bound state will be directly linked to the relative rate of gene expression. The riboswitch operates in the equilibrium state of ligand binding. An example for a riboswitch under thermodynamic control is the *Vibrio vulnificus add*-riboswitch (Rieder et al., 2007).

However, for other riboswitches the kinetics of ligand binding controls ligand binding. For the case of regulation by transcription termination the decision between ON or OFF has to be made before the transcribing RNA polymerase has progressed past the start codon of the regulated gene. This limited time-window for genetic regulation can result in riboswitch behavior predominantly being dependent on the on-rate of ligand binding. Examples of riboswitches that operate under kinetic control include the *B. subtilis pbuE* and the *xpt-pbuX* purine riboswitches (Gilbert et al., 2006b; Wickiser et al., 2005; Lemay et al., 2006).

1.3.3 Distribution of riboswitches in different taxonomic groups

Prokaryotic riboswitches

Riboswitches have been found in a broad variety of different prokaryotic organisms of very different taxonomic groups, reviewed in Barrick and Breaker 2007. Some riboswitches like the SAM-I and the thiamine pyrophosphate (TPP) riboswitch are found in most taxonomic groups within prokaryotes, while others like the purine and the SAM-II riboswitch have been found only in a few groups. The abundance of riboswitch control in different taxonomic groups varies greatly. Firmicutes and Fusobacteria display a high number of riboswitches for many different ligands in their genome, while for example in Chlamydia and Spirochetes only few riboswitches for one or two ligands have been discovered (Barrick and Breaker, 2007).

Which mechanism is adopted by a particular riboswitch also differs greatly within species. For example the SAM-I riboswitch acts mainly by transcription attenuation in Firmicutes while in Bacterioidetes and γ-proteobacteria it exclusively utilizes translation attenuation. It seems however, that within different taxonomic groups one or the other mechanism largely predominates. For example, the γ-proteobacteria almost exclusively utilize translation attenuation in riboswitch-dependent gene regulation while in firmicutes riboswitches using transcription attenuation predominate (Barrick and Breaker, 2007).

Eukaryotic riboswitches

The overwhelming majority of riboswitches identified to date were found in prokaryotic organisms. However, a few riboswitches have been found in eukaryotes. The first eukaryotic riboswitches identified were the TPP riboswitches in plants, for example *Arabidopsis thaliana* and filamentous fungi such as *Neurospora crassa* (Sudarsan et al., 2003a). The structure of the *A. thaliana* TPP riboswitch

has been solved (Thore et al., 2008). TPP riboswitches have also been identified in green algae (Croft et al., 2007). In addition to the TPP riboswitch, cytokinin binding riboswitches have been proposed in *A. thaliana* (Grojean and Downes, 2010). The regulation mechanisms of eukaryotic riboswitches can include altered splicing of the mRNA, which is not relevant for prokaryotic riboswitches.

1.3.4 The purine riboswitch

The class of purine riboswitches can be divided in two subclasses, the adenine and the guanine riboswitches. However, these two subclasses share an extremely similar aptamer domain, whose specificity for either adenine or guanine is determined by a single nucleotide. The purine riboswitch aptamer domain consists of about 70 nucleotides, which upon ligand binding assumes a structurally well conserved fold based on a three way junction (Fig. 1.4 A). This aptamer domain was first crystallized in 2004 by two groups independently, with a resolution of approximately 2 Å for both the adenine binding and the guanine binding variant (Serganov et al., 2004; Batey et al., 2004).

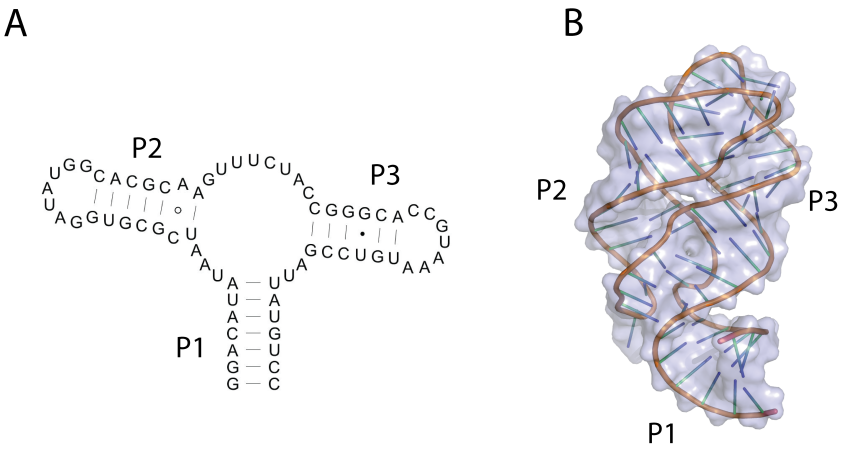


Figure 1.4: The structure of the purine riboswitch. A: Sequence and secondary structure diagram B: Crystal structure of the purine riboswitch (PDB-file 2g9c) shown as a cartoon representation.

The crystallographic data confirmed the secondary structure of three helices ar-

ranged around a central three-way junction (Fig. 1.4 B). The 5′ and 3′ end of the aptamer domain form the P1 helix which branches into the three-way junction. The three dimensional structure shows coaxial stacking of the P1 and P3 helices, while the P2 and P3 helices are extending in the same general direction interacting via their terminal loops. The structures confirmed that the ligand specificity for either adenine or guanine is determined by the pyrimidine 74. This nucleotide makes a Watson-Crick-type base pair with the bound purine ligand, thus specifically binding adenine in case of U74 and guanine in case of C74. The ligand is bound in a central binding pocket hidden from bulk solvent (Fig. 3.1 B). Held in place by six hydrogen bonds, binding is exceptionally strong for a molecule this size (only ten non-hydrogen atoms in the case of adenine). Adenine contains five nitrogen atoms all of which form hydrogen bond contacts with the aptamer, and the N6-amino group forms two hydrogen bonds, one with U74 and one with a water molecule located near the only opening of the binding pocket towards bulk solvent (Serganov et al., 2004). In addition to small nucleobase-like ligands, it has been shown that the purine riboswitch can also accommodate nucleoside ligands like 2'-deoxyguanosine (Edwards and Batey, 2009). Structural data show that the binding pocket is enlarged via movement of the A47 and U48 (Fig. 1.5). This implies a degree of flexibility for this riboswitch and supports a binding model of induced fit upon ligand binding.

It is very important to note that up to date no structure of the purine riboswitch without a purine in the binding pocket has been determined. All crystal structures contain an exogenous ligand or, in one case one nucleotide flips its base into the binding pocket thus mimicking a ligand-bound state (Delfosse et al., 2010).

The folding of the purine riboswitch aptamer domain has been studied extensively (Lemay et al., 2006; Lin and Thirumalai, 2008; Lee et al., 2010; Leipply and Draper, 2011). FRET experiments (Lemay et al., 2006) revealed the presence of three distinct states in solution prior to ligand binding. The first state represents the unfolded structure with just the helices formed with no tertiary structure. The

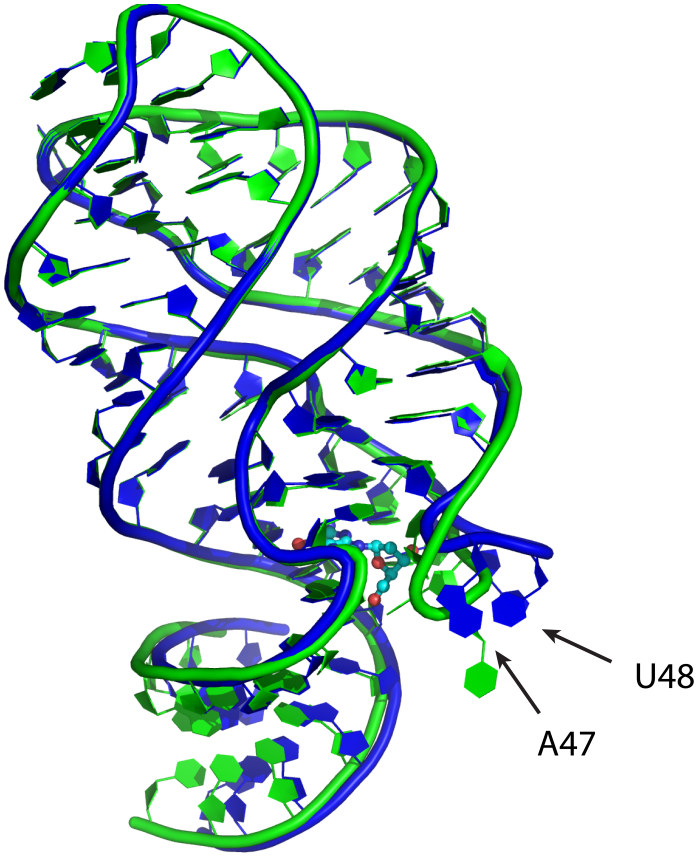


Figure 1.5: Structure of the purine riboswitch (PDB-file 3d2s, blue) bound to 2'-deoxyguanosine (cyan carbon atoms) superpositioned with the 2,4,6-triaminopyrimidine bound structure (PDB-file 2g9c, green). The nucleotides A47 and U48 which move away from the core to generate the space needed for the sugar moiety of the ligand are highlighted with arrows.

second state, the intermediate state has P1 and P3 stacked but no interaction with P2, while the third state, the folded state has P1 and P3 stacked and the loop-loop interaction between P2 and P3 formed. The ligand can bind to the folded structure, thereby stabilizing it, resulting in the formation of a stable pseudoknot conformation of the three-way junction.

1.3.5 The SAM-I riboswitch

Seven different riboswitches have been identified for the ligand SAM (Winkler et al., 2003; Corbino et al., 2005; Lu et al., 2008; Weinberg et al., 2008; Poiata et al., 2009). The SAM riboswitches contain different aptamer domains that

are structurally diverse. Crystal structures determined for SAM-I, SAM-II and SAM-III illustrate this (Fig. 1.6). The binding mode of the SAM and the ligand recognition patterns vary significantly between the different classes (Montange and Batey, 2006; Gilbert et al., 2008; Lu et al., 2008).

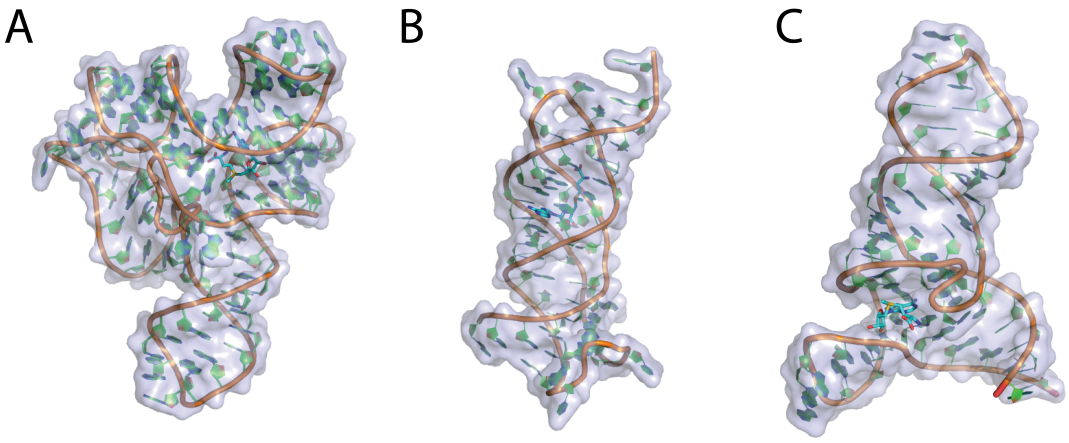


Figure 1.6: The diversity of SAM-binding aptamer domains. The structures of the SAM-I, SAM-II and SAM-III have been solved to date. A: Structure of the SAM-I riboswitch (PDB-file 2gis), B: Structure of the SAM-II riboswitch (PDB-file 2qwy), C: Structure of the SAM-III riboswitch (PDB-file 3e5c). All riboswitches are displayed in a cartoon representation. The SAM ligand is shown with cyan carbon atoms.

The structural diversity of SAM-aptamers is also reflected in the different sizes of the aptamer domains from the various classes of SAM-riboswitches. While the SAM-I aptamer domain contains approximately 100-150 nucleotides SAM-II and SAM-III are significantly smaller with only 40-65 nucleotides (Fig. 1.6). Over the whole class SAM-recognizing aptamers can differ in size by up to a factor of five (Wang and Breaker, 2008).

The SAM-I riboswitch was originally discovered in gram-positive bacteria, especially *B. subtilis*, as a domain of conserved sequence, called the S-box (Grundy and Henkin, 1998), even though riboswitches were unknown at the time and the regulation of gene expression was hypothesized to depend on "some unknown factor" (Grundy and Henkin, 1998). The S-box motif was found eleven times in the *B. subtilis* genome regulating eleven operons comprising 25 genes in total.

These genes are involved in sulfur metabolism and the biosynthesis of methionine and SAM (Tomsic et al., 2008). The description of the S-box motif as a riboswitch, which then became the SAM riboswitch was reported in 2003 (Winkler et al., 2003). Subsequent discovery of unrelated riboswitches that also bind and recognize SAM led to renaming this the SAM-I riboswitch. The natural ligand of this riboswitch is SAM, a central metabolite in prokaryotes as well as eukaryotes (Loenen, 2006; Grillo and Colombatto, 2008).

SAM-based gene regulation operates by transcription termination, which occurs in the ligand-bound state making the SAM-I riboswitch an OFF-switch (Winkler et al., 2003). The ligand specificity of the aptamer domain has been investigated (Winkler et al., 2003; Lim et al., 2006; Montange et al., 2010). SAM has a submicromolar binding affinity (5nM for the *B. subtilis yitJ* riboswitch) and the chemically very similar S-adenosylhomocysteine (SAH) typically has an over 100 fold higher dissociation constant (Winkler et al., 2003; Montange et al., 2010). The secondary structure of the SAM-I riboswitch is organized by a central four-way junction from which all four helices branch (Fig. 1.7 A). Similar to the purine riboswitch the ends form the P1 helix and the other helices P2-P4 (5' to 3') are stem loops (Montange and Batey, 2006) while helix P1 is unterminated.

However, P2 and P3 are not simple stem loops but contain several unpaired nucleotides. In the case of P2 the helix is interrupted by a k-turn motif (see below and Fig. 1.7 A) and the distal part of the helix is called P2a. The P3 helix contains a bulge located at its center.

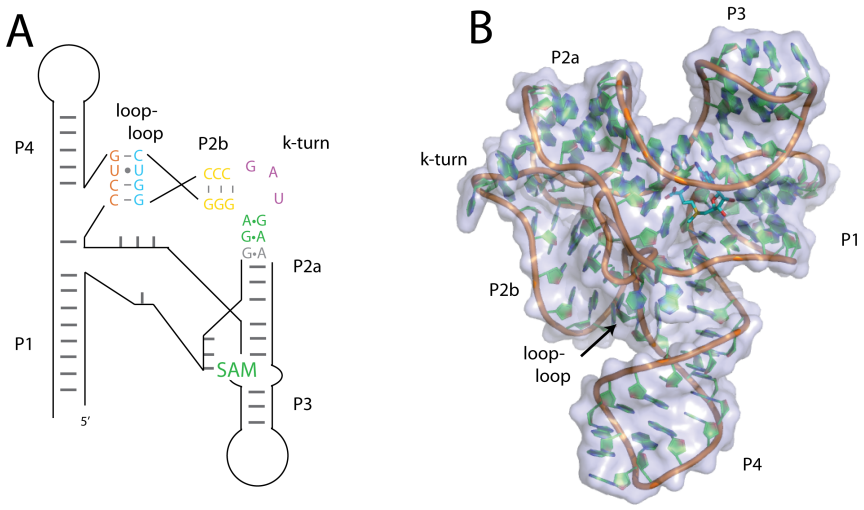


Figure 1.7: Structure overview for the SAM-I riboswitch. A: Schematic representation of the secondary structure. B: Cartoon representation of the SAM-I riboswitch. The SAM-ligand is displayed in cyan carbon atoms.

A crystal structure was first determined for the *Thermoanaerobacter tencongesis* SAM-I riboswitch (Montange and Batey, 2006). The SAM binding pocket is located in the core of the folded aptamer (Fig. 1.7 B), the SAM ligand assumes a highly compact conformation with the methionine moiety stacking back onto the adenine moiety (Fig. 1.8). The ligand interacts with the binding pocket via multiple hydrogen bonds tethering each part of the ligand to the pocket (Fig. 1.8). The adenine moiety is bound by A45 and U57, which are prevented from forming a standard Watson-Crick base pair. A45 and U57 form hydrogen bonds with the N1, N6 and N7 nitrogen atom of the SAM, thus utilizing both the Hoogsteen and the Watson-Crick edges. The ribose moiety of the SAM forms two hydrogen bonds, one from the 2'-hydroxyl group to the O4'-ring oxygen of the C47 ribose and one from the 3'-hydroxyl group to the U7 O4'-ring oxygen. The methionine moiety is held in place via hydrogen bonds from the carboxyl group to the Watson-Crick edge of G11 and the N2 amino group of G58. Also the methionine amino group of the SAM ligand hydrogen bonds to the 2'-hydroxyl group and N3 of G58. These multiple hydrogen bonds together with a good shape complementarity of the ligand and the binding pocket result in a sub-

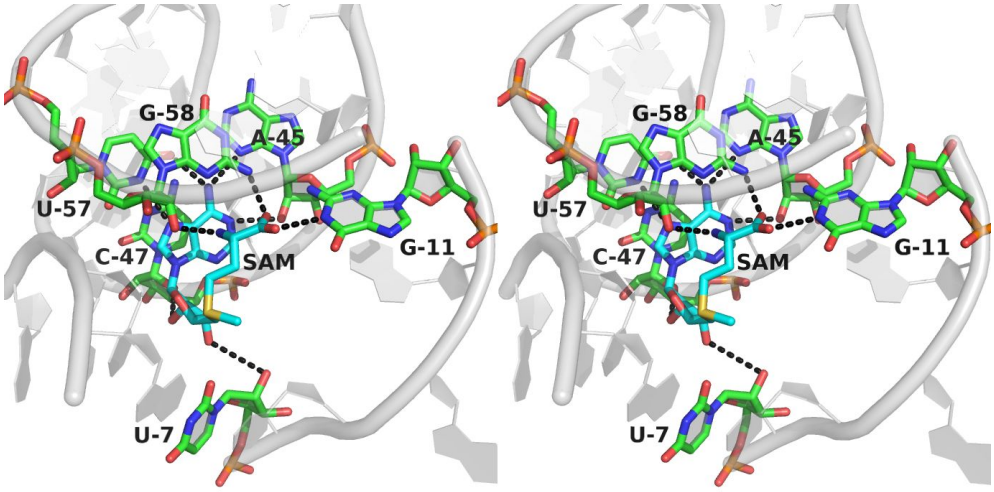


Figure 1.8: Wall-eyed stereo view of the SAM ligand (cyan carbon atoms) bound to the SAM-I riboswitch (grey cartoon). The nucleotides interacting with the ligand are shown with green carbon atoms and the hydrogen bonds between ligand and riboswitch are highlighted in black dashed lines. The figure illustrates the tight interaction of all parts of the SAM ligand with the riboswitch nucleotides.

micromolar (approximately 300 nM) affinity of SAM for the *T. tengcongensis* SAM-I riboswitch.

The SAM-I riboswitch was also crystallized in the absence of ligand (Stoddard et al., 2010). The apo structure is highly similar to the SAM-bound structure and the most prominent difference is the conformation of A46. This nucleotide is located outside of the binding pocket in the presence of the SAM ligand but flips into the binding pocket in the absence of ligand. The A46 in the binding pocket forms hydrogen bonds to both A45 and U57 (Fig. 1.9).

1.4 k-turns

As mentioned above, the SAM-I riboswitch contains a kink-turn (k-turn) motif in the P2 helix. The k-turn motif was first described about 10 years ago (Klein et al., 2001) as a repeated structural RNA motif found in ribosomal crystal structures (Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000). The k-turn is a conserved intra-helical three-dimensional structural motif, that introduces a kink

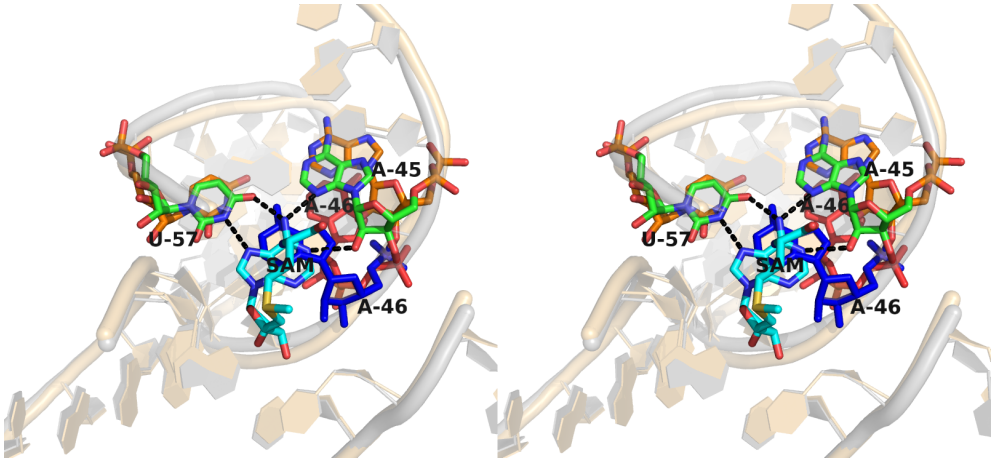


Figure 1.9: Superposition of the SAM-bound structure (PDB-file 2gis) and the apo-structure of the SAM-I riboswitch. The SAM-bound structure is displayed in grey cartoon representation and the apostructure in beige cartoon representation. The key nucleotides and the ligand are shown as sticks. The SAM ligand is shown in cyan carbon atoms, A45 & U57 in green carbon atoms for the SAM-bound structure and in amber carbon atoms for the apo structure. The A46 is shown in red for the SAM-bound structure and in dark blue for the apo structure.

into an RNA helix with an approximate included angle of 60° between the helix arms. The nomenclature is used as in (Liu and Lilley, 2007) and a schematic is displayed in Figure 1.10. It can be divided into three parts. The central part consists of 2-4 unpaired bases in one strand which is termed the bulged strand (b-strand). These nucleotides are named 5' to 3' as L1 to Ln in case of an n-base bulge. The helix to the 5' side of the bulge is termed the canonical helix (C-helix), since it consists of standard Watson-Crick base pairs. To the other side of the bulge the non-canonical helix (NC-helix) contains 2-4 non-standard base pairs before becoming a helix of standard base pairs. It is the bulge region and the NC-helix that mainly form the characteristic interactions in k-turns.

The structural data on multiple examples of the k-turn motif facilitated an analysis of the interactions stabilizing the k-turn fold (Fig. 1.11) (Turner and Lilley, 2008). The two base pairs adjacent to the bulge are conserved A-G base pairs with G1b and A2b pairing to A1n and G2n respectively in a *trans* Hoogsteen-sugar edge interaction. Another conserved interaction is a hydrogen bond between the L1-2' hydroxyl and the A1n-N1, which has been shown to be essential for k-turn

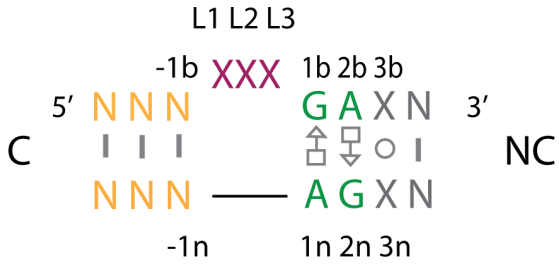


Figure 1.10: Schematic representation of a k-turn with the nomenclature for the nucleotides as used here.

folding (Liu and Lilley, 2007). Also the 2'-hydroxyl group of L2 and L3 form hydrogen bonds to the non-bridging *proS* phosphate oxygens in their proximity. After their first description, k-turns have been found in multiple RNA species of different types and have been collected in a web-based database (Schroeder et al., 2010). These RNAs include the 23S and 16S rRNA (Klein et al., 2001), riboswitches (Montange and Batey, 2006; Blouin and Lafontaine, 2007), snoRNAs (Cléry et al., 2007; Li and Ye, 2006), mRNAs in addition to riboswitches (Chao and Williamson, 2004) and snRNAs (Woźniak et al., 2005). All of the aforementioned RNAs have structurally confirmed examples but there are many more examples of putative k-turns that have not been verified yet. k-turns appear to be an important structural feature of ribosomes (Klein et al., 2001), and are recognized by a number of k-turn binding proteins, including L7Ae (Hamma and Ferré-D'Amaré, 2004), S7, S11 (Dunkle et al., 2011) and L24.

k-turns in a simple duplex adopt the tightly kinked conformation only in the presence of metal ions (Goody et al., 2004; Matsumura et al., 2003). The best-studied example is the *Haloarcula marismortui* Kt-7 from the 23S ribosomal subunit and this k-turn has been used as a model system to characterize the motif. When an RNA duplex containing Kt-7 and and labeled with fluorescein and Cy3 is titrated with metal ions, magnesium ions induce folding of Kt-7 with $[Mg^{2+}]_{\frac{1}{2}} = 80\mu M$ while sodium ions induce folding at much higher concentrations as reflected in $[Na^{+}]_{\frac{1}{2}} = 30mM$ (Liu and Lilley, 2007). Folding isotherms are well fitted as a two state process with a Hill coefficient of $n_{Hill} \approx 1$. These data

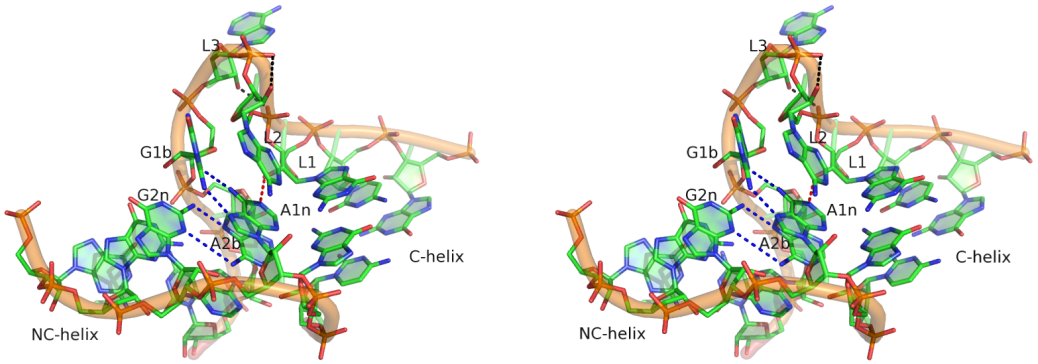


Figure 1.11: Wall-eyed stereo view of the *H. marismortui* Kt-7 showing the key interactions for k-turn formation. The hydrogen bonds of the two G-A pairs are shown in blue, the crucial hydrogen bond from the O2' of L1 to N1 of A1n in red and the hydrogen bonds form the O2' of L2 and L3 to the neighboring non-bridging *proS* oxygens are shown in red.

indicate that metal ions act as diffuse counter ions that neutralize the phosphate charge to allow close approach of the backbone in the folded state, rather than requiring metal ion binding at specific sites. Furthermore, it was shown that k-turns in isolation can be folded by k-turn binding proteins in the absence of metal ions (Turner et al., 2005; Woźniak et al., 2005). The ubiquity of k-turn motifs in key biological RNAs emphasizes the importance of an understanding of k-turn structure and function on a molecular level. The binding of k-turns to their cognate proteins is likely to be an important event in ribosome biogenesis.

1.5 Objectives

1.5.1 Development and validation of a method for RNA ligand docking

In order to advance the field of RNA-ligand docking, a method was developed and tested using a suitable model system. The *Bacillus subtilis xpt-pbuX* guanine riboswitch mutant C74U (GRA) was selected as a model system because of its computational and experimental tractability. Additionally this system offers a

wealth of published binding and structural data. The binding pocket of the purine riboswitch is small, completely buried from bulk solvent and does not contain any formal charge. These characteristics are similar to those of model binding pockets used for method development and validation in the protein docking field (Wei et al., 2002; Rosenfeld et al., 2003; Brenk et al., 2006). These features were expected to facilitate interpretation of the results and comparison to the protein-ligand docking field. Also the experimental amenability of the system would enable validation of computational results and thus provide a solid base for later generalization. A comprehensive investigation of the performance of the chosen computational approach both by retrospective and prospective methods was attempted. To date this is the first study of RNA-ligand docking to include high-resolution structures of predicted RNA-ligand complexes.

The high solubility of adenine compared to guanine makes the adenine-responsive C74U purine riboswitch a suitable system for the exploration of RNA-ligand docking.

1.5.2 Virtual ligand screening of the SAM-I riboswitch

The SAM-I riboswitch occurs with very similar sequence multiple times in the genome of gram positive bacteria such as *B. subtilis, B. anthracis* and *B. cereus*, regulating essential metabolic pathways. This makes this riboswitch an ideal target for drug design, occurring both in pathogenic and non-pathogenic organisms. Any compound binding to the SAM-I riboswitch with high affinity and thereby perturbing gene expression would affect multiple protein targets. A virtual ligand screening was performed to identify new ligands.

1.5.3 Study of the sequence dependence of k-turn folding in context of the SAM-I riboswitch

The SAM-I riboswitch contains a k-turn in its P2 helix, it is relatively easy to crystallize and SAM-binding is an indicator of folding. Therefore, this riboswitch offers an opportunity to study k-turn folding in an RNA with long range tertiary interactions. The sequence-dependent folding of the k-turn from this riboswitch was investigated in isolation and in context of the SAM-I riboswitch in order to elucidate the contribution of tertiary interactions to k-turn folding. Data obtained from folding studies using FRET measurements and calorimetry were backed up with structural evidence from X-ray crystallography.

Chapter 2

Materials and Methods

2.1 Virtual ligand screening / molecular docking

2.1.1 Receptor preparation

The coordinates of the receptor molecules used, PDB files 2g9c (Gilbert et al., 2006a) for the GRA and 2gis (Montange and Batey, 2006) for the SAM-I riboswitch, were obtained from the PDB (www.pdb.org). Hydrogen atoms were added using the Sybyl software (Tripos). The positions of these hydrogens were subsequently minimized using the MAB force field (Gerber and Müller, 1995) as implemented in Moloc (Gerber molecular design) while keeping all non-hydrogen atoms rigid. Afterwards all atoms not part of the final receptor setup (all atoms not part of the RNA and potential water molecules kept for docking) were removed. The ligand binding site was defined via a sphere set, which was calculated according to the principles described below. This sphere set was based on a cubic grid and the coordinates of the original ligand. The sphere set serves two purposes during docking:

1. it defines the binding site and is used to determine the area within the bind-

ing site with a reduced dielectric constant for calculating the electrostatic potential

2. during docking ligand atoms are matched with the spheres to generate potential binding modes

Therefore, different sphere sets can be used during receptor preparation and docking.

For docking of GRA the same sphere set was used for receptor preparation and docking. The distance between neighboring grid points was ≈1.5 Å and the grid extended 2 Å around the original ligand. For the SAM-I riboswitch docking, two different sphere sets were used. The sphere set used for receptor preparation and the docking of the SAM-analogues extended 5 Å around the original SAM-ligand and neighboring grid points were ≈ 2Å apart. For docking, a smaller sphere set, 1.5 Å around the adenine moiety of the SAM ligand with neighboring grid points ≈ 2 Å apart, was used. Partial charges for all receptor atoms were obtained from the AMBER99 force field parametrization (Wang et al., 2000). The charges of the phosphate oxygen atoms were modified to obtain a resulting net charge of zero per nucleotide, thus mimicking charge screening by diffuse counter ions (Detering and Varani, 2004; Moitessier et al., 2006). The electrostatic potential of the receptor was calculated using DelPhi (Nicholls and Honig, 1991), with an internal dielectric constant of 2 and an external dielectric constant of 78. The van der Waals energy map was calculated using the DOCK utility CHEMGRID and the excluded volume map using the DOCK utility SOLVMAP (Mysinger and Shoichet, 2010).

2.1.2 Sphere set generation

Originally, DOCK 3.5.54 and the accompanying scripts provided a sphere set consisting of the atoms of the original ligand in the crystal structure. In case of no original ligand being present, a substitute had to be provided manually. This

procedure introduces a bias towards the correct binding mode for the original ligand and ligands based on the scaffold of the original ligand. To avoid this problem and to generate sphere sets that define the binding site well, an algorithm was developed that generates a regular lattice of spheres with the lattice axes derived from the original ligand geometry. From this lattice of spheres, only spheres around the original ligand and not too close to any receptor atoms were included in the final sphere set for docking (Fig. 2.1).

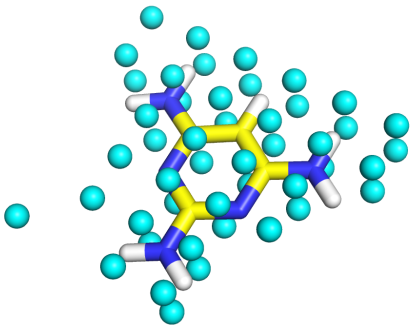


Figure 2.1: Example sphere set generated for docking. Yellow carbon atoms: ligand from the PDB-file used for docking (2,4,6-triaminopyrimidine). Cyan: Spheres generated based on that ligand. The spheres are arranged on a simple cubic crystal lattice. The lattice orientation was based on the ligand geometry, resulting in one lattice vector being orthogonal to the ligand plane. Only spheres not clashing with any RNA atoms were kept, so the sphere set reflects the shape of the binding site.

2.1.3 Small molecule preparation

From the in-house databank of commercially available compounds with precalculated physicochemical properties (Brenk et al., 2008), the screening sets were assembled by filtering for compounds with desirable properties (as described in the corresponding results chapters). The training set for the GRA and the SAM analogues for the SAM-I riboswitch were generated manually in SMILES format (Weininger, 1988). The protocol for the preparation of small molecules for docking was similar to the one described in (Mpamhanga et al., 2009).

Protonation states and tautomeric states were calculated using in-house scripts based on the OEToolkit (Openeye, Santa Fe, NM). Three-dimensional coordinates and low energy conformations were calculated using OMEGA2 (Openeye, Santa Fe, NM). Partial charges and desolvation energy were calculated using AMSOL (http://comp.chem.umn.edu/amsol/; Wei et al. 2002; Brenk et al. 2006). Finally, molecules were aligned on their ring systems and stored in a hierarchical format (Wei et al., 2002).

2.1.4 Molecular docking

After receptor and small molecules preparation docking was performed using DOCK 3.5.54 (Lorber and Shoichet, 1998; Wei et al., 2002; Mysinger and Shoichet, 2010). Sampling parameters were:

- receptor and ligand bins were set to 0.5 Å
- overlap bins were 0.4 Å
- distance tolerance for ligand atoms and receptor matching sites ranging from 1.1 Å to 1.2 Å

Only docking poses not placing ligand atoms in the area occupied by the receptor were scored for electrostatic interaction energy, van der Waals energy and desolvation energy. Only the best-scoring representation in terms of docking pose, as well as protonation and tautomerization states, was stored in the resulting output files.

2.1.5 Docking analysis, pharmacophore filtering and molecular graphics

Docking results were visualized and analyzed with the Sybyl software (Tripos). Pharmacophore filtering was also carried out using the Sybyl interface. The

pharmacophore was built in Sybyl by placing hydrogen bond acceptor and donor functionalities at the desired locations. The filtering was performed using the UNITY module (Tripos). All diagrams of molecular structures were generated in PyMOL (http://pymol.org). Two-dimensional chemical structures were generated using Symyx Draw 3.3 (Symyx solutions).

2.2 RNA synthesis and purification

2.2.1 Synthesis of fluorescently labeled RNAs

The RNA oligonucleotides were synthesized via *t*-BDMS phosphoramidite chemistry (Beaucage and Caruthers, 1981) as described previously (Wilson et al., 2001). Fluorescein and Cy3-conjugated oligonucleotides were attached at the 5'-termini as phosphoramidites during synthesis. The oligoribonucleotides were deprotected in 25% ethanol / ammonia solution at 65 °C for 1h and dried. Subsequently, oligoribonucleotides were dissolved in 300 μl 1M tetrabutylammonium fluoride in tetrahydrofuran to remove the *t*-BDMS protecting groups and agitated at 20 °C for 16h prior to desalting using a G25 Sephadex column (NAP columns, Pharmacia) and ethanol precipitation. All oligonucleotides were purified by denaturing polyacrylamide gel electrophoreses and recovered from gel fragments by electroelution or diffusion followed by an ethanol precipitation. The fluorescently-labeled RNA was purified by reversed-phase HPLC on a C18 column (ACE10, Advanced Chromatography Technologies) using an acetonitrile gradient with an aqueous phase of 100mM triethylammonium acetate (pH 7.0). The duplex species were annealed in 50 mM Tris-HCl (pH 7.5), 25 mM NaCl by slow cooling from 90 °C to 4 °C and purified by electrophoresis.

2.2.2 *In vitro* transcription of riboswitch aptamer domains

RNA for both riboswitches was transcribed *in vitro* from templates generated by PCR. The primers used for the PCR were as follows, written from 5' to 3':

- generic 5' primer: GCGCGCGAATTCTAATACGACTCACTATAG
- 3' primer for SAM-I riboswitch: CGGCTCATCTTTCAACGTTTCCGCT
- 3' primer for the GRA: TGGACATAATCGGACATTTACGGTGCCCGG

The final PCR mixture (typically 1ml final volume) contained: 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100, 0.2 mM each dNTPs, 1 μM each primer, 600 $\frac{ng}{ml}$ template plasmid and 1 $\frac{U}{ml}$ Taq DNA polymerase (prepared in house). The plasmids containing the riboswitch sequences (containing the GRA and the *T. tengcongensis* SAM-I U34C/A94G sequences) were generously supplied by Robert T. Batey. PCR was then performed in a thermocycler using the following regime: 5min 94 °C initial denaturation, then 25 cycles of 30 sec 94 °C denaturation, 30 sec 55 °C annealing and 30 sec 72 °C extension, followed by a final 7 min elongation step at 72 °C.

The transcription reaction itself (typically 12.5 ml final volume) contained 40 mM Tris-HCl (pH 8.0), 4 mM each NTP (Sigma), 32 mM $MgCl_2$, 10 mM DTT, 3.2 mM spermidine, 0.01 % Triton X-100, 0.25 $\frac{mg}{ml}$ T7 RNA polymerase (prepared in-house), 8.0 % (v/v) DNA template from the PCR reaction described above and 0.1 $\frac{U}{ml}$ inorganic pyrophosphatase (Sigma). Transcriptions were allowed to proceed for 3h at 37 ° C and precipitated with 3 volumes of cold ethanol. Typically transcriptions were of 12.5 ml volume, and the template PCR accordingly of 1 ml volume.

The precipitated transcription reactions were subsequently purified by denaturing polyacrylamide gel electrophoresis. The gels contained 12 % polyarcyrlamide (19:1), 90 mM Tris borate (pH 8.3), 10 mM EDTA and 7 M urea. The RNA was

visualised by UV shadowing and the product band was cut out. The RNA was then electro-eluted in a 3000 Da molecular weight cut-off dialysis tube against 90 mM Tris borate (pH 8.3) and 10 mM EDTA overnight at 4 °C. RNA was then concentrated using 3000 Da molecular weight cut-off spin concentrators and quantified by UV absorbance. The molar extinction coefficients used were $\epsilon_{260nm,GRA} = 680000 \frac{1}{M\ cm}$ and $\epsilon_{260nm,SAM-I} = 850000 \frac{1}{M\ cm}$ for the GRA and the SAM-I riboswitch, respectively.

2.3 Binding assays

2.3.1 Fluorescence binding assay for the GRA

The fluorescence spectra for the binding assay were recorded on an SLM-Aminco 8100 flourimeter. The sample buffer consisted of 50 mM Tris HCl (pH 8.3), 100 mM KCl and 10 mM $MgCl_2$. The spectra of the fluorophore 2-aminopurine were collected from 330 to 450 nm with an excitation wavelength of 300 nm and corrected for lamp fluctuations and instrumental variations. 2-aminopurine fluorescence was determined as the integral of the collected spectra.

The binding constant of the fluorophore (compound **3** / 2-aminopurine, Tab. 3.2) was determined first. For this RNA was titrated into a solution containing 200 nM 2-aminopurine in sample buffer. The free 2-aminopurine concentration was calculated and fitted to a simple two state binding model, resulting in a binding constant of $K_L = 140$ nM.

Competition experiments were carried out by titrating ligand into a solution of fluorophore and GRA. Initially, a 2-aminopurine spectrum (c=213 nM) was collected prior to adding GRA to a final concentration of 243 nM RNA and 200 nM fluorophore. After RNA addition, the solution was left to equilibrate for 10 min and a spectrum was recorded. Subsequently, competition ligand was incrementally added to a final concentration of at least 1.2 mM (except for compound **25**

where the maximum concentration was 0. 68mM due to limited solubility). After each addition, the sample was left to equilibrate for 3 min prior to measurement of a spectrum. This time was estimated to be sufficient for equilibration based on reported data on the folding kinetics for the purine riboswitch (Buck et al., 2007) Data were processed and analyzed in Excel (Microsoft). Free 2-aminopurine concentration was fitted to the following binding model with two complexes and dilution effects were taken into account (Yan et al., 2005):

$$R^3 + aR^2 + bR - K_L K_D R_{tot} = 0 \tag{2.1}$$

where

$$a = L_{tot} + X_{tot} + K_L + K_D - R_{tot} \tag{2.2}$$

$$b = K_D L_{tot} + K_L X_{tot} + K_L K_D - R_{tot}(K_L + K_D) \tag{2.3}$$

In these equations, R is the concentration of unbound RNA, R_{tot} the total RNA concentration, L_{tot} the total 2-aminopurine concentration, X_{tot} the total concentration of test compound (competition ligand), K_L the binding constant of 2-aminopurine, and K_D the binding constant of the test compound. All parameters except R and K_D are known at the start of the experiment. R is obtained from the fluorescence reading using the known K_L. K_D can then be obtained from minimizing

$$\Phi = \sum \chi_i \tag{2.4}$$

where, for the ith data point,

$$\chi_i = R^3 + aR^2 + bR - K_L K_D R_{tot} \tag{2.5}$$

The obtained binding constants were averaged from three measurements and rounded to a precision of 0.01 mM (0.001 mM for compound **5**).

2.3.2 ITC measurements for the SAM-I riboswitch with varied k-turn sequences

ITC measurements for the SAM-I riboswitch with varied k-turn sequences were carried out at 30 °C using the MicroCal VP-ITC instrument (GE Healthcare). The protocol was similar to that described in (Montange et al., 2010). The SAM-I sequence was GGCUUAUCAAGAGAGGUGGAGGGACUGGCCCGAC-GAAACCCGGCAACCAGAAAUGGUGCCAAUUCCUGCAGCGGAAACGUU GAAAGAUGAGCCG, which carries the U34C/A94U mutation and is defined here as the natural sequence. Modifications in the k-turn were inserted into a plasmid bearing this sequence by site-directed mutagenesis. During the titration, a solution of 100 μM SAM in sample buffer was injected into the cell containing 10 μM SAM-I riboswitch in sample buffer. The sample buffer contained 40 mM K-HEPES (pH 7.5), 100 mM KCl and 10 mM $MgCl_2$. Following an initial 1 μl injection 35 injections of 8 μl were performed. The resulting calorimetric data were analyzed using the MicroCal Origin software and fitted to a single-site binding model. Each titration was performed in triplicate and the obtained fitting parameters were averaged. Free energy and the dissociation constant for SAM were calculated from:

$$\Delta H^\circ - T\Delta S^\circ = \Delta G^\circ = -RTlnK_D \tag{2.6}$$

where R is the gas constant and T is the temperature in Kelvin.

2.3.3 Competition ITC measurements for the SAM-I riboswitch

The competition ITC measurements to determine binding of the screening compounds to the SAM-I riboswitch were conducted using the VP-ITC calorimeter from Microcal. Sample buffer was prepared by adding 2 %(v/v) of 100 mM com-

pound in DMSO to the basic ITC buffer described above (40 mM K-HEPES (pH 7.5), 100 mM KCl, 10 mM $MgCl_2$), resulting in a final concentration of 2 mM for the competition ligand. For the reference measurement in the absence of competition ligand DMSO only was used. RNA was buffer-exchanged by centrifugation five times in a microconcentrator and rediluting with sample buffer, obtaining a final concentration of about 200 μM. RNA was then diluted down to a final concentration of 5 μM and inserted into the cell. RNA was quantified prior to buffer exchange, since the presence of compound did not allow accurate RNA quantification due to overlapping absorbance. The SAM solution for the syringe was prepared by adding 10 μl of 10 mM SAM solution to 2 ml of sample buffer containing the competition ligand. For the titration, 35 8 μl injections were performed following an initial 1 μl injection. The raw data were analyzed using the MicroCal Origin software to obtain the heat evolved for each injection. These data were then fitted in Excel to a single state binding model derived from the model used in the original software. The fitting function supplied by the software assumes the following model:

$$\Delta H = 0.5 V_{cell} \Delta H^{\circ} \left(1 + \frac{1 - \frac{X_t}{M_t} - \frac{n}{K_a M_t}}{\sqrt{(1 + \frac{X_t}{M_t} - \frac{n}{K_a M_t})^2 - 4 M_t}} \right) \tag{2.7}$$

where V_{cell} is the volume of the cell, K_a the apparent association constant of SAM, X_t the concentration of SAM in the cell and M_t the concentration of RNA in the cell. This model requires $\Delta H \approx 0$ for large $\frac{X_t}{M_t}$ (towards the end of the titration), however due to minute differences in buffer composition between cell and syringe, there is a residual ΔH_{buff}. Therefore, the model was modified to account for this:

$$\Delta H = 0.5 V_{cell} \Delta H^{\circ} \left(1 + \frac{1 - \frac{X_t}{M_t} - \frac{n}{K_a M_t}}{\sqrt{(1 + \frac{X_t}{M_t} - \frac{n}{K_a M_t})^2 - 4 M_t}} \right) + \Delta H_{buff} \tag{2.8}$$

All titrations could be fitted to this modified model. Due to high requirement of RNA and ligand, titrations were carried out only once per ligand. A binding constant for the competition ligand was calculated from the apparent binding constant of SAM in the absence and presence of competition ligand using the following formula (Zhang and Zhang, 1998):

$$K_D^{complig} = \frac{c_{lig}}{\left(\frac{K_D^{SAM+lig}}{K_D^{SAM}} - 1\right)} \tag{2.9}$$

2.3.4 FRET folding assay

Absorption spectra were measured in 90mM Tris-borate (pH 8.3) in 120 μl volumes using a NanoDrop 2000C spectrophotometer (Thermo Scientific). Spectra were deconvoluted using a corresponding RNA species labeled only with Cy3. Fluorophore absorption ratios were calculated using a MATLAB program. Fluorescence spectra were recorded in 90 mM Tris-borate (pH 8.3) at 4 °C using an SLM-Aminco 8100 fluorimeter. Spectra were corrected for lamp fluctuations and instrumental variations and polarization artifacts were avoided by setting the excitation and emission polarizers crossed at 54.7 °. Values of FRET efficiency were measured using the acceptor normalization method (Clegg et al., 1992) implemented in MATLAB. FRET efficiency as a function of metal ion concentration was fitted to a two state model for ion-induced folding:

$$E_{FRET} = E_0 + \Delta E_{FRET}\left(\frac{K_a[M]^n}{1 + K_a[M]^n}\right) \tag{2.10}$$

where E_0 is the FRET efficiency of the RNA in the absence of added metal ions, ΔE_{FRET} the increase in FRET efficiency between start and saturation. [M] is the Mg^{2+} ion concentration, K_a the apparent association constant for metal ion binding and n the Hill coefficient. Data were fitted in KaleidaGraph to this equation by nonlinear regression and the metal concentration at which the

transition was half complete is given by:

$$[M]_{\frac{1}{2}} = \left(\frac{1}{K_a}\right)^{\frac{1}{n}} \tag{2.11}$$

2.4 Crystallography and structure determination

2.4.1 Crystallization

All crystals were grown using the hanging drop method in 24-well plates on VDX slides (Hampton Research). RNA was buffer-exchanged and concentrated to give a 400 μM solution in sample buffer (10 mM K-HEPES (pH 7.5) for the GRA and 40 mM Na-cacodylate (pH 7.0) for the SAM-I riboswitch) with ligand present. Crystals were then grown in drops generated by mixing 1 μl of RNA solution with 1 μl of mother liquor. Crystal drops were then seeded using the microseeding technique. Seed stocks were obtained from crystals grown unseeded. For the GRA, these were grown with 2,6-diaminopurine as the ligand, for the SAM-I riboswitch the natural sequence was readily crystallized with SAM. Obtained crystals from three drops were then combined with 50 μl of mother liquor and mixed thoroughly. All crystal trays were stored at $\approx$ 20 °C.

Crystallization of GRA-complexes

The sample buffer for the GRA contained 1mM of the respective ligand. The mother liquor for the conditions yielding the crystal for data collection contained:

1. 2-aminopurine complex (compound **3**, Tab. 3.2): 320 mM ammonium acetate, 12 mM cobalt hexammine, 17.5 % PEG 3000, 10 mM K-HEPES (pH 7.5)

2. N6-methyladenine complex (compound **15**, Tab. 3.2): 440 mM ammonium

acetate, 12 mM cobalt hexammine, 20.0 % PEG 3000, 10 mM K-HEPES (pH 7.5)

3. complex with compound **24** (Tab. 3.3): 560 mM ammonium acetate, 12 mM cobalt hexammine, 22.5 % PEG 4000, 10 mM K-HEPES (pH 7.5)

4. complex with compound **26** (Tab. 3.3): 800 mM ammonium acetate, 12 mM cobalt hexammine, 25.0 % PEG 4000, 10 mM K-HEPES (pH 7.5)

5. complex with compound **27** (Tab. 3.3): 800 mM ammonium acetate, 12 mM cobalt hexammine, 25.0 % PEG 3000, 10 mM K-HEPES (pH 7.5)

Crystallization of the SAM-I riboswitch

The sample buffer contained 1mM SAM. The mother liquor for the crystals used for data collection contained:

1. G2nA variant: 40 mM Na-cacodylate (pH 7.0), 16 mM spermine-HCl, 10 mM $MgCl_2$, 80 mM KCl, 14 %(v/v) MPD

2. G2nU/G3nU variant: 40 mM Na-cacodylate (pH 7.0), 12 mM spermine-HCl, 5 mM $BaCl_2$, 80 mM KCl, 12 %(v/v) MPD

2.4.2 Data collection, processing and structure modeling

For the structure of GRA, in complex with compound **15**, data were collected at the Diamond Light Source beamline I03 (Didcot, UK). For all other structures of GRA data were collected at beamline ID14-1 at the ESRF (Grenoble, France). Data were indexed, integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). Structures were solved using MOLREP (Vagin and Teplyakov, 1997) from the CCP4 suite of programs (CCP4, 1994) using PDB-file 2g9c as a starting model. Refinement was performed using REFMAC5 (Vagin et al., 2004) from the CCP4 suite of programs. Model building was performed in COOT (Emsley and

Cowtan, 2004; Emsley et al., 2010). Ligand topology files were generated using the PRODRG webserver (Schüttelkopf and van Aalten, 2004). All coordinates, except for the 2-aminopurine complex, were deposited in the protein data bank with accession codes 2xo0,2xo1,2xnw and 2xnz.

For the SAM-I riboswitch, the data on the G2nA variant were collected at beamline BM14 at the ESRF (Grenoble, France). Data were indexed, integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). For the G2nU/G3nU variant, data were collected at beamline ID14-4 at the ESRF (Grenoble, France). Data were indexed and scaled using MOSFLM/Scala (Leslie, 2006; Battye et al., 2011) from the CCP4 suite of programs. Both structures were solved using MOLREP (Vagin and Teplyakov, 1997) and refinement was carried out in REFMAC5 (Vagin et al., 2004). Model building was performed in COOT (Emsley and Cowtan, 2004; Emsley et al., 2010). Composite omit maps were calculated using Autobuild in PHENIX (Adams et al., 2010). Finally structures were deposited in the PDB (accession codes 2ygh and 2ydh).

Chapter 3

Virtual screening for ligands for the *xpt-pbuX* guanine riboswitch C74U mutant from *B. subtilis*

3.1 Overview

To develop and validate a method for RNA-ligand docking, a model system was chosen. The DOCK 3.5.54 software was adapted to perform the RNA-ligand docking and the approach was validated by comparison of docking results for known ligands and non-binders with the published data. Subsequently, a large set of compounds was screened to identify potential novel binders. These candidate compounds were then tested experimentally and for a subset of the confirmed binding compounds the binding mode was determined by X-ray crystallography. The results are discussed with respect to the implications for RNA-docking and molecular docking in general.

3.2 Results

3.2.1 Screening of a test set of known binders and non-binders

Application of DOCK 3.5.54 to RNA targets

To advance the field of RNA-ligand docking the GRA was chosen as a test system. At the outset of the study, DOCK 3.5.54 had only been applied to protein targets and DOCK 3.5.54 was lacking the necessary parameter files for the monomeric subunits of RNA. The required parameters were atom type and partial charge for each atom. The atom types were assigned manually and the partial charges were obtained from the AMBER99 force field (Wang et al., 2000). RNA naturally is a poly-anion in solution, which requires counter ions to maintain local electro-neutrality (as reviewed in Draper 2004). To reflect this, the partial charges for the phosphate oxygens were adjusted so that for each monomeric subunit the sum of partial charges was zero, as it was reported in previous studies (Moitessier et al., 2006; Detering and Varani, 2004).

To probe the setup described above, a test set of compounds with known binding properties was assembled from the literature. This test set was to include both binders and non-binders and comprised all compounds with published binding properties for that receptor at this time. The data set was assembled from two publications (Mandal and Breaker, 2004; Gilbert et al., 2006a). Gilbert et al. 2006a used GRA while Mandal and Breaker 2004 used the *V. vulnificus* adenine riboswitch (AR) as their test system. These two riboswitches are virtually identical with respect to their their overall fold and their binding pocket (Fig. 3.1). The overall overlay of GRA and AR has an RMSD of only 1.45 Å (based on all atoms) and all binding site residues are conserved and in a similar conformation. The key binding site nucleotides (A21, U22, U47, U51, A52, U74 & U75) overlayed within 0.24 Å based on all residue atoms. Therefore it was assumed that

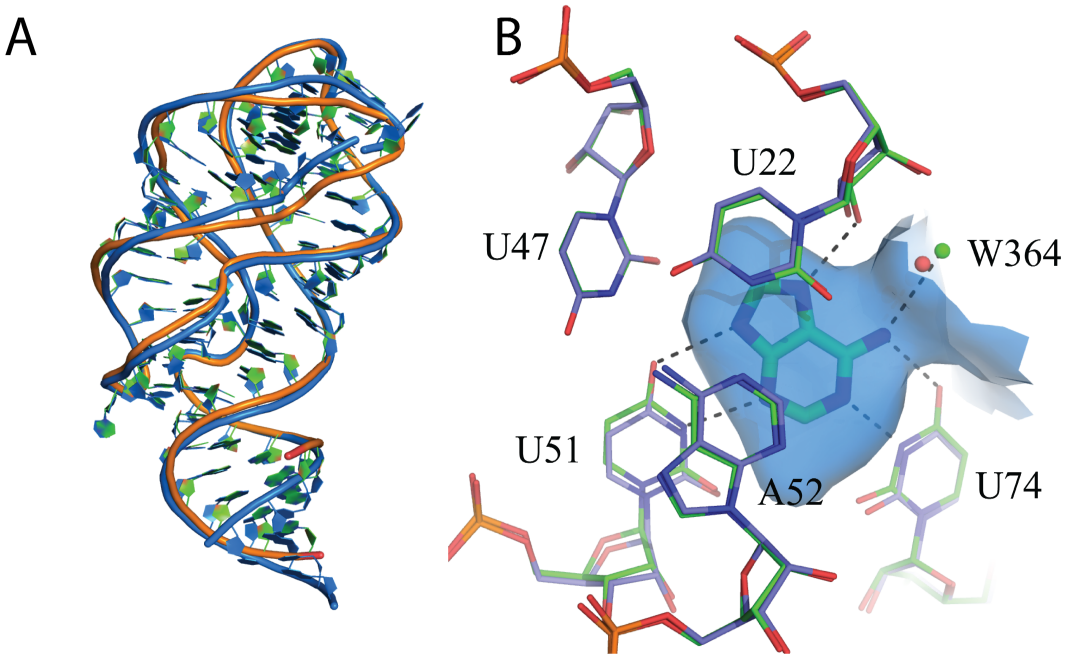


Figure 3.1: Comparison of the *B. subtilis xpt-pbuX* guanine riboswitch C74U mutant (GRA) and the *V. vulnificus* adenine riboswitch (AR) A: Superposition of GRA (amber/green) and AR (blue). The overall fold is virtually identical with an RMSD of 1.45 Å. B: Comparison of the binding sites from both crystal structures with the AR with green carbon atoms and a green water molecule, GRA with blue carbon atoms with a red water molecule. The solvent accessible surface of the AR binding site is displayed in blue. The position of all bases in contact with the ligand is conserved.

binding data was transferable between both riboswitches.

From the compounds included in this test set (Tab. 3.2), three were cocrystallised with a riboswitch and the coordinates were available from the PDB. These compounds were 2,6-diaminopurine (Gilbert et al. 2006b; PDB-file 2b57), 2,4,6-triaminopyrimidine (Gilbert et al. 2006a; PDB-file 2g9c) and adenine (Serganov et al. 2004; PDB-file 1y26). The structures for 2,6-diaminopurine and 2,4,6-triaminopyrimidine were in complex with GRA while the structure for adenine was in complex with AR.

For docking, the structure of the 2,4,6-triamino-pyrimidine in complex with GRA was selected, since at that time, this was the highest resolution crystal structure available with a resolution of 1.7 Å (Gilbert et al., 2006a).

In order to define the binding site for docking, DOCK 3.5.54 requires a set of

dummy atoms, sometimes also referred to as "spheres". During docking, ring systems of the compound are used as anchor points and ring atoms are placed in the vicinity of the docking spheres. The compound-target complexes resulting from this procedure are then scored (Shoichet et al., 1992). Accordingly, the sphere set is vital to successful docking. Therefore a systematic protocol for the generation of the sphere set was developed (Chapter 2.1.2). The binding modes of the cocrystallised compounds and the separation of scores for the known binders and non-binders in a score-ranked table were used to assess the performance of the docking procedure. In the initial docking runs no water molecules were present in the receptor. With the so prepared receptor the known ligands scored better on average than the known non-binders. Also the binding modes of 2,6-diaminopurine and 2,4,6-triamino-pyrimidine were predicted correctly. However, for adenine an incorrect binding mode was predicted. In the crystallographically observed binding mode the adenine ligand forms a Watson-Crick base pair with U74, the N6-amino group pointing in the gap between U22 and U74 (Fig. 3.1 B; Serganov et al. 2004). The predicted binding mode was oriented differently, so that the Watson-Crick edges of U74 and the adenine ligand were still facing each other but the N6-amino group now pointed in the gap between U51 and U74 and the Hoogsteen edge instead of the sugar edge of the adenine ligand now faced U51.

The importance of water molecule 364

Close analysis of the available crystal structures led to the hypothesis, that the water molecule 364 located between U74 and U22 (Fig. 3.1 B) is important for the orientation of adenine in the binding pocket. A water molecule is located in this position in all three published cocrystal structures available at that time. In the adenine bound structure (1y26) this water is part of the hydration sphere of magnesium ion 125, which is absent in the structure used for docking since no magnesium was present in the crystallization conditions. However, adenine is

the only ligand for which this was expected to influence the ligand orientation. Both 2-6-diaminopurine (compound **1**) and 2,4,6-triamino-pyrimidine(compound **5**) contain several amino-groups occupying both possible locations for the adenine N6-amino group, while adenine only contains one amino-group. Therefore water molecule 364 was included into the receptor and kept present in the docking run. The effect of the inclusion of the water molecule was threefold (Tab 3.2): Firstly, adenine was predicted with the correct binding mode and secondly the separation between binders and non-binders in the score-ranked list of docked compounds improved. Thirdly, N6-methyladenine (compound **5**), a known binder, now scored worse and the predicted binding mode did not agree with the expected binding mode. This was no surprise since the N6-methyl group was expected to occupy the same location as water molecule 364.

To probe the binding mode prediction of the docking approach for a ligand without a known crystal structure the GRA structure complexed with 2-aminopurine (compound **3**) was determined. To further investigate the effect of water molecule 364, the crystal structures of GRA in complex with N6-methyladenine (compound **15**) was determined (Tab. 3.1). Crystallization of the GRA in complex with N-methyladenine was carried out by Colin M. Hammond, a rotation student in the lab, under supervision of the author. The expected binding mode and the docking results for compound **15** with and without the water molecule were compared to these crystal structures. The overall fold of GRA in complex with compound **3** was virtually identical to that of the riboswitch bound to compound **5** which was used for docking (Fig. 3.2 A). The overall RMSD between the GRA structures complexed with compounds **3** and **5** based on all atoms was 0.16 Å. Compound **3** bound to the riboswitch as predicted by the docking algorithm. The purine ring of the 2-aminopurine was in the same orientation as found in the case of the adenine complex. The 2-aminogroup pointed into the gap between C74 and U51 (Fig. 3.2 B). It is noteworthy that the water molecule was still observed in the crystal structure (Fig. 3.2) even though there is no amino-group of the ligand in

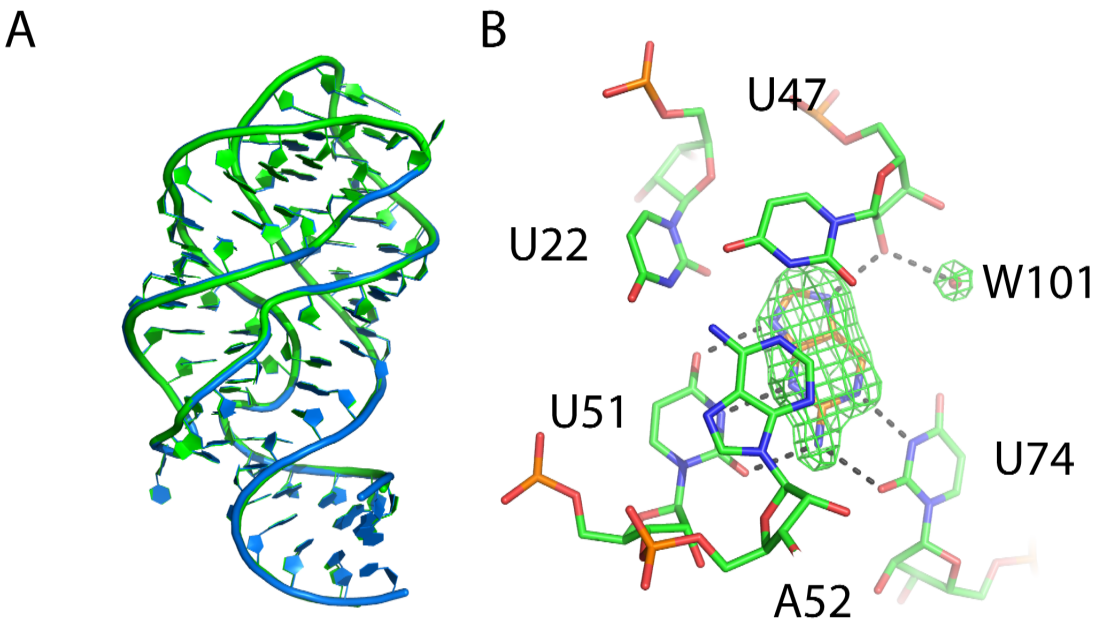


Figure 3.2: Crystal structure of 2-aminopurine bound to GRA at 1.8 Å. A: superposition of 2-aminopurine bound structure (blue) with the structure which was used for docking (green). The overall fold is very well conserved and can be regarded as identical within margin of error (RMSD 0.16 Å based on all atoms). B: binding mode of 2-aminopurine in the binding pocket. The riboswitch binding pocket is displayed in green carbon atoms, and the 2-aminopurine in amber carbon atoms. The green mesh represents the $F_o - F_c$ electron density at 3.5 σ calculated without the ligand and water molecule 101 present. The ligand binding mode can be assigned unambiguously from these data. In red water molecule W101 is shown, which is found in a location very similar to W364 in the structure used for docking.

close range to form a hydrogen bond with. The water molecule 101 only formed a hydrogen bond with the O2' hydroxyl group of U22. In case of the co-crystal structure with N6-methyladenine the overall fold and the binding pocket were conserved as well. As expected, the orientation of the purine ring in the binding pocket was the same as for the adenine ligand (Fig. 3.3). The N6-methyl group displaced the water molecule .

The protonation state of 8-azaadenine

The high score for 8-azaadenine (compound **6**) was not attributable to an irrelevant binding site conformation. However, it was hypothesized that a negatively

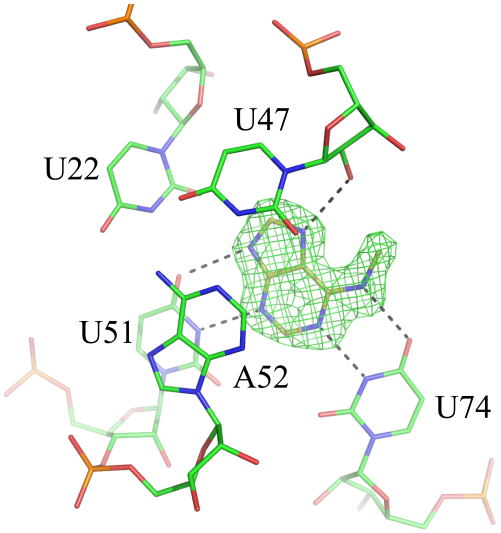


Figure 3.3: Crystallographically determined binding mode of N6-methyladenine (amber carbon atoms) in complex with GRA (green carbon atoms). The green mesh shows the $F_o - F_c$ electron density map contoured at 3σ, calculated by omitting the ligand from the final model. The water molecule present in the other crystal structures is displaced by the N6-methyl group of the ligand.

charged deprotonated species might be dominant under the experimental conditions where the binding experiments were carried out. The pKa of the hydrogen atom bearing N7 was calculated to be 7.6. Accordingly the compound would be predominantly deprotonated under assay conditions (pH 8.3). The negatively charged, deprotonated species when tested by docking received a score of $15.54 \frac{kJ}{mol}$ which would result in rank 22 for this compound.

Correcting for the 8-azaadenine protonation state and using the docking score of N6-methyladenine without the water present would result in a perfect separation of ligands and decoys in the test set.

3.2.2 Screening of a large compound library and prediction of novel ligands

In order to further explore the ability of the chosen approach for RNA-ligand docking, a virtual screening library was compiled, docked into the receptor and top-scoring compounds were selected for experimental evaluation. Our in-house

virtual compound data-base containing approximately 6.5 million unique compounds was filtered for compounds fulfilling the following criteria:

1. 18 or less non-hydrogen atoms
2. 1 or more hydrogen bond donor and acceptor group
3. total charge between -1 and +2
4. 1 or 2 ring systems

These criteria ensured that the candidate compounds do not violate fundamental requirements of the binding site.

After docking of these compounds all those that have a positive van der Waals score were removed from the core-ranked list prior to further analysis. This filter step removed compounds that were too large to fit the binding pocket or did not match the binding site shape and ensured that the statistical analysis is not skewed by obviously non-fitting molecules (Verdonk et al., 2004). This resulted in a final set of 2592 compounds. To asses the performance of the docking algorithm the enrichment of the known ligands and decoys from the test among the top-scoring compounds in the large data set was displayed in a ROC-plot (receiver operating characteristic, Fig. 3.4).

The enrichment for the ligands is very strong, which is reflected in an area under curve (AUC) of 0.98. With a theoretical maximum of 1 and a value of 0.5 for a random distribution of the ligands within the ranked list this corresponds to a near-perfect performance. The decoys were enriched among the top scoring compounds as well. The AUC for the decoy set was 0.75, which is still well above random distribution.

3.2.3 Experimental validation of the screening results

Encouraged by the above mentioned results that validate the *in silico* approach, five compounds of unknown binding characteristics from the top-ranking com-

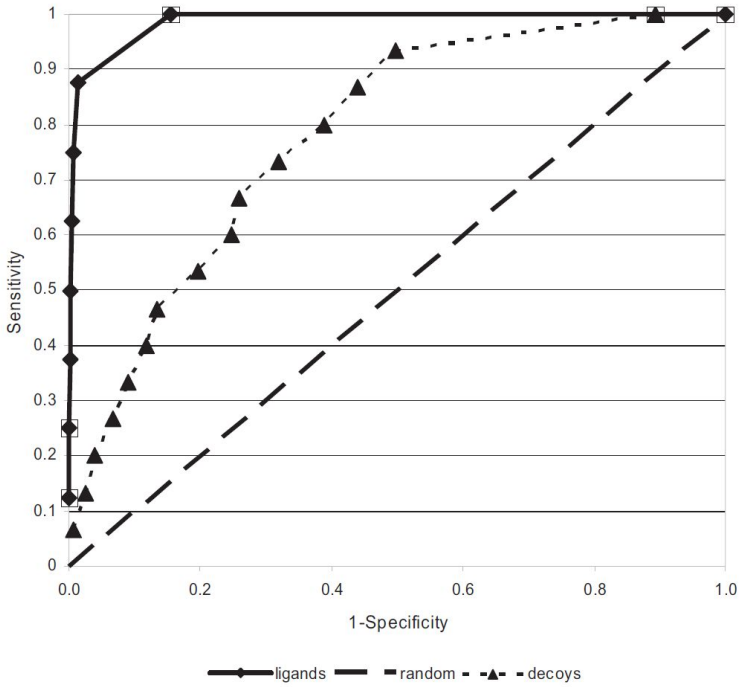


Figure 3.4: ROC plot of docking results for known ligands and decoys mixed into in a large set of compounds with unknown binding properties. The sensitivity (fraction of known compounds) was plotted against 1-specificity (fraction of unassigned compounds in data set). An area under curve (AUC) of 0.98 for the known ligands (solid line) and 0.75 for the decoys (dotted line) was obtained. For comparison, a random prediction (dashed line) would yield an AUC of 0.5.

pounds of the virtual screening set were selected for experimental characterization after close inspection of their predicted binding mode (Tab. 3.3). From these five compounds **25**, **26** and **27** closely resembled known binders while **24** and **28** were based on molecular scaffolds not previously known from ligands of this riboswitch.

In order to confirm the binding of the selected compounds to the riboswitch, a fluorescence competition assay based on 2-aminopurine binding was used. Upon riboswitch binding the stacking of the 2-aminopurine ligand leads to a drop in quantum yield for the 2-aminopurine fluorescence (Mulhbacher and Lafontaine, 2007). The ligand to be tested was then titrated into a cuvette containing 2-aminopurine and the riboswitch and binding was observed as an increase of flu-

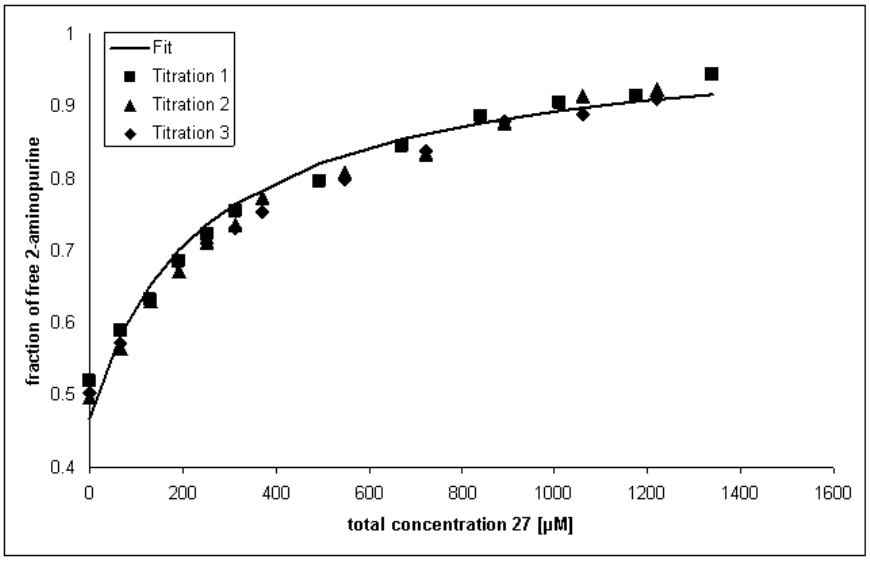


Figure 3.5: Plot of fraction of free 2-aminopurine against the concentration of compound **27**. Data were fitted to a one binding site model with two equilibria. Data from three titrations are shown together with the resulting fit (solid line).

orescence resulting from the displacement of 2-aminopurine by the tested ligand in the riboswitch binding pocket. Representative data are shown in Fig. 3.5 using the data obtained for compound **27** as displacement ligand. From the five selected compounds all except compound **25** were confirmed to bind to GRA. Interestingly, compound **25** is based on the triazine scaffold (Tab 3.3) which is also present in compounds **2** (Tab 3.2), **26** and **27** (Tab 3.3).

The determined affinities for the four binding compounds ranged from 80 μM to 650 μM which corresponds to approximately one-thousandfold weaker binding than the natural ligand adenine. The affinities for the compounds based on scaffolds not previously observed in ligands for this riboswitch (compounds **24** and **28**) are considerably weaker (370 μM and 650 μM respectively).

For the four confirmed ligands cocrystallisation in complex with the riboswitch was attempted and a crystal structure was obtained for three of these four ligands. The resolutions of the obtained crystal structures ranged from 1.5 to 1.7 Å (Tab. 3.1). For compound **28** no crystal structure could be determined. Small crystals could be grown with this ligand but no crystals of sufficient size for data collection

were obtained. The limiting factor seemed to be the solubility of the ligand in the crystallization buffer which limited the maximum concentration of ligand in the crystallization conditions. This results in combination with the weak affinity of the ligand for the riboswitch results in a relatively low fraction of the RNA being bound to the ligand within the crystallization drop. However, the small crystals observed are a further indication that the compound is a true ligand since until now no ligand-free structure of this riboswitch is known and crystals can only be obtained in presence of a ligand.

The overall fold and the binding pocket conformation of the riboswitch cocrystallised with compounds **26** is virtually identical compared with the triaminopyrimidine bound structure (RMSD of 0.16 Å based on all atoms) which was used for docking. The ligand orientation could not be unambiguously determined from the observed density alone, due to ligand symmetry (Fig. 3.6). Therefore, the position of the water molecule corresponding to water molecule 364 in the structure complexed with compound **5** was taken into account. This water was shifted away from the ligand by 0.75 Å, which was indicative of the methyl group pointing in this direction. The ligand forms hydrogen bonds with the Watson Crick edges of both U51 and U74.

Additionally a water molecule was now located in the complex interacting with both a ring-nitrogen of the ligand and the 2'-hydroxyl group of U22. No corresponding water molecule was observed in the structure complexed with compound **5**. The binding mode of compound **26** was different from the predicted binding mode by the docking algorithm, which had the methyl group facing the O4 of U51. However when docking was performed with the shifted water molecule the correct binding mode was predicted with an RMSD of predicted to observed binding mode of 0.29 Å.

The crystal structure of compound **27** bound to GRA was somewhat surprising. Again, fold and conformation of the binding pocket were conserved between this structure and the structure used for docking. Compound **27** however, adopted

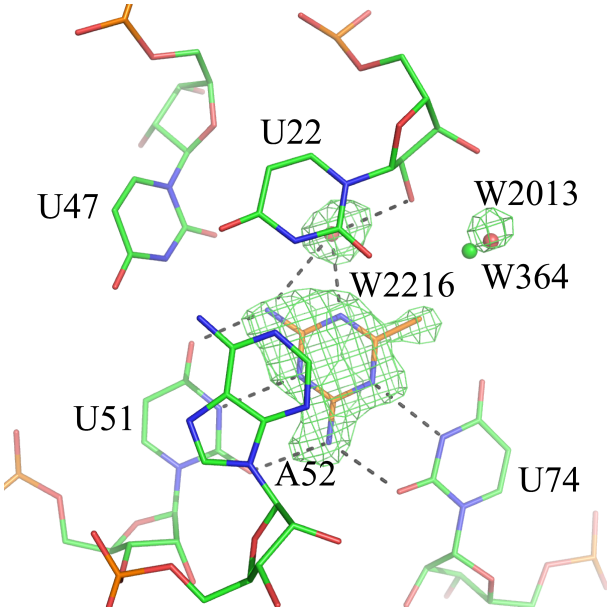


Figure 3.6: Crystallographically determined binding mode of compound **26** (amber carbon atoms) to GRA(green carbon atoms). The green mesh shows the $F_o - F_c$ electron density map contoured at $3\,\sigma$, calculated by omitting the ligand from the final structure. The water molecule W2013 at the opening of the binding site of this structure is shown in red, it is displaced by 0.75 Å compared to W364 from the structure used for docking (shown in green). An additional water molecule (W2216) is found in the binding pocket which hydrogen bonds to the ring nitrogen and one of the exocyclic amino groups of the ligand and the O2' hydroxyl group of U22.

two binding modes in the GRA binding site (Fig. 3.7). In each of these orientations the ring moiety was located in the center of the binding pocket with the plane being identical to that of compound **5** (Tab. 3.2). Also one of the two exocyclic amino groups pointed into the cleft between U74 and U51 in both cases. The other amino group pointed towards the O4 of U51 in one conformation and towards the O4 of U74 in the other conformation. The ratio of the occupancies for these binding modes was estimated to 70%:30% (amino group towards O4 of U51 and U74 respctively) based on crystallographic B-factors. This corresponded to a difference in binding energy of $\Delta G \approx 2\ \frac{kJ}{mol}$ as followed from the Boltzmann distribution.

Initially only the second binding mode was predicted by the docking algorithm with an RMSD of 0.23 Å. The protocol only reported the single binding mode

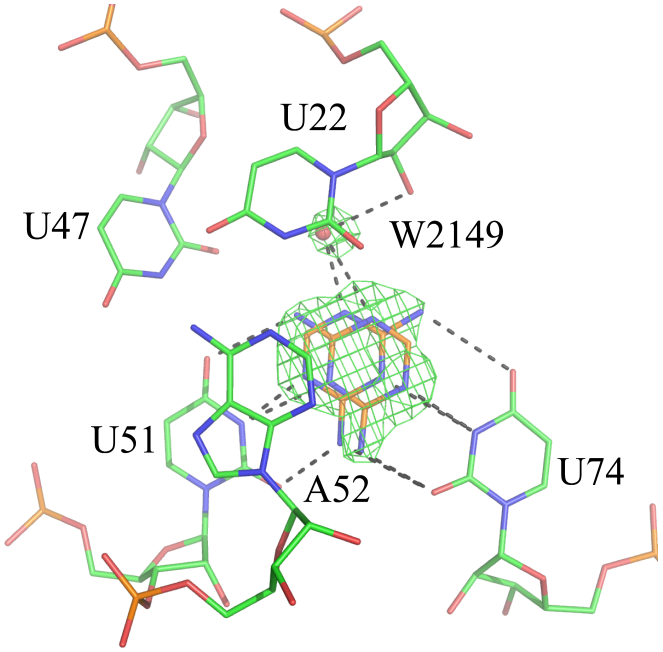


Figure 3.7: Crystallographically determined binding mode of compound **27** (amber carbon atoms) to GRA (green carbon atoms). The green mesh shows the $F_o - F_c$ electron density map contoured at 2.5σ, calculated by omitting the ligand from the final model. A water molecule (W2149) is found in a similar location as W2216 in the **26**-GRA complex. The ligand adopts two distinct binding modes.

which is scored best, but when compound **27** was docked into the GRA binding site while allowing multiple binding modes, the first binding mode was also predicted with a similar docking score differing only by 1.5 $\frac{kJ}{mol}$.

The crystal structure of compound **24** with the riboswitch again showed conservation of the overall fold and the binding pocket conformation. The binding mode of the ligand could be assigned without any ambiguity due to the lack of symmetry within the ligand and the well defined electron density for the ligand in this 1.5 Å structure (Fig. 3.8). The ligand hydrogen bonded with the Watson-Crick edges of U51 and U74, the 2'-hydroxyl of U22 and water molecule W2329 which was located in an equivalent position to water molecule 364 of the structure complexed with compound **5** used for docking. The binding mode observed in the crystal structure agreed with the binding mode predicted by docking with an RMSD of 0.21 Å which was within experimental error.

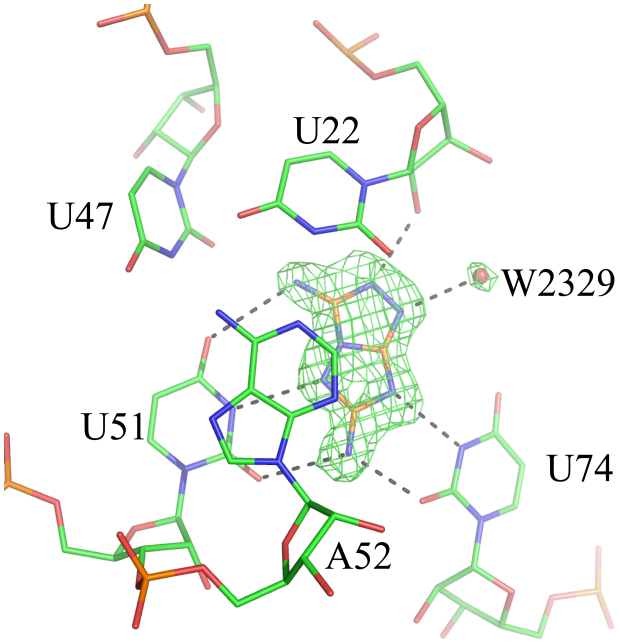


Figure 3.8: Crystallographically determined binding mode of compound **24** (amber carbon atoms) to GRA (green carbon atoms). The green mesh shows the $F_o - F_c$ electron density map contoured at 2.5 σ, calculated by omitting the ligand from the final model. The binding mode of the ligand can be unambiguously assigned from these data. Water molecule W2329 is found in a similar position as W364 in the structure used for docking.

When overlaying the binding mode of compound **5** with compound **24** (Fig. 3.9), it became apparent that the molecular interaction pattern for both ligands is very similar. The guanidino-like moiety including the extra cyclic amino group pointing into the cleft between U51 and U74 was present in both ligands. Also both ligands had an amino group interacting with the O4 of U51, which added up to three hydrogen bonds between U51 and each ligand. The interaction of the ligands with U74 is different. While compound **5** formed three hydrogen bonds with U74, compound **24** only formed 2 hydrogen bonds with this base. But compound **24** formed an extra hydrogen bond to the 2'-hydroxyl group of U22 which had no corresponding interaction in the binding mode for compound **5**.

3.3 Discussion

3.3.1 Setup of DOCK 3.5.54 for RNA-ligand docking

Docking algorithms have been applied to proteins for a long time now and drug discovery has benefited from this technique (Ripphausen et al., 2010). However, RNA has remained mostly unexplored as a target, even though various RNAs are validated drug targets (Fulle and Gohlke, 2010). Studies reporting predictions of new binders which are subsequently tested by structural and biochemical assays are lacking. However, these studies are of great importance when probing the strengths and weaknesses of a docking protocol (Kolb et al., 2009). Partly this dearth of studies is due to the polyanionic nature of RNA polymers which poses a more complex problem to docking than proteins which are not systematically charged. Even though individual amino acids in a protein can carry a charge, the peptide backbone is electroneutral and both positively and negatively charged side chains are found in proteins. In contrast each nucleotide in an RNA molecule carries a negative charge associated with the under physiological conditions deprotonated phosphate group. To fulfill the fundamental requirement of electro neutrality in solution this is balanced by counter ions (Draper, 2004, 2008). These can be either bound in specific binding sites or act as diffuse counter ions which are close to the phosphate backbone but not localized to a specific position. In essence the concept of a diffuse counter ion is equivalent to a slightly elevated concentration of positively charged ions in the region around the negatively charged backbone where the ions are otherwise virtually unrestrained in mobility. Most docking algorithms require a precise location for each atom to be considered in docking, which poses a challenge to specify the RNA target for docking since diffuse counter ions are not easily parameterized in such a setup. If the effects of counter ions are not included into the docking setup the resulting electrostatic potential is modeled incorrectly and docking results would be strongly biased towards positively charged compounds.

The approach for the results described above was to artificially elevate the partial charges of the phosphate oxygens bearing the negative charge to maintain overall neutrality. This emulates the effect of diffuse counter ions, whose charge is shielding the charged groups on the RNA backbone so it contributes less to the electrostatic interaction with the ligand. This approach has been used in earlier attempts of RNA-ligand docking (Detering and Varani, 2004; Moitessier et al., 2006). Apart from the treatment of this electrostatic asymmetry, the scoring function is another crucial factor in docking. The scoring function of DOCK 3.5.54 used here is based on three distinct terms:

1. electrostatic energy
2. van der Waals energy
3. desolvation energy

The reasons that led to the choice of this scoring function are as follows: Firstly, it is physics based and requires only few parameters specific for RNA. The generation of lookup files for partial charges and atom types for the RNA-atoms were sufficient for application of the scoring function. Atom types were easily assigned manually and the partial charges taken from the AMBER force field, which has been tested for RNA molecules in the past (Deng and Cieplak, 2010). Secondly, with the score as the sum of three distinct terms each of which has a physical meaning, modification or balancing of the scoring function is possible. The scoring function of DOCK 3.5.54 was developed and tested for proteins. In principle, the physics based scoring function should work outside the protein setting equally well, since it does not contain any protein specific adaptations. Because RNA-ligand interactions are known to be dominated by polar contacts (Hermann, 2000), it was expected that the scoring function would require some scaling of the individual terms to account for the RNA situation. Surprisingly, no modifications were necessary to achieve discrimination between ligands and

decoys in the test set (Tab. 3.2). Given the dominance of charged interactions in RNA-ligand complexes compared to protein-ligand complexes, it is actually somewhat surprising that the same balance of terms as in protein ligand was observed to perform at a comparable level. However, from a fundamental view the same balance of terms should be applicable in any situation for a physics-based scoring function like the one used. The dominance of electrostatic interactions should simply be reflected in the average magnitude of partial charges. In that respect the docking results confirm the validity of the used scoring function.

The GRA was chosen as a model system because the binding pocket resembles model binding sites used in protein ligand docking for validation of docking protocols (Wei et al., 2002; Brenk et al., 2006). The GRA binding site is small and buried from bulk solvent like the mentioned protein model binding sites and the known ligands are small and nearly rigid. These characteristics helped rationalize why certain predictions contradicted the experimental evidence. It was then possible to adjust the parameters of the docking to achieve agreement between prediction and experiment. This is discussed below in the context of the retrospective docking results.

3.3.2 Retrospective results

The observed separation of known ligands and decoys in the test set was nearly perfect. This is remarkable, especially since the decoys in the test set were very similar to the ligands (Tab. 3.2). All compounds in the test set were based on either modified purine or pyrimide scaffolds. All decoys were planar and able to form at least some of the interactions formed by known similar ligands. This indicates a sensitivity of the docking algorithm for minor changes that are experimentally known to prohibit binding.

For an ideal scoring function the docking score and the binding energy (related to the binding affinity) should be correlated strongly. This correlation is weak

for the known ligands of the purine riboswitch (Tab. 3.2). Considering the many approximations and that only three factors of binding energy are accounted for, this is not overly surprising. The lack of this correlation is also common for protein ligand docking (Kim and Skolnick, 2008) and there is no reason why RNA-ligand docking should not suffer from this phenomenon.

The separation of the ligands and decoys is not perfect (Tab. 3.2) and deserves a more detailed discussion. Compound **15** was initially assigned an unfavorable score when docking into the GRA binding site with the water molecule 364 present. From the crystal structure obtained this could easily be rationalized and the obtained score for docking without the water molecule present was much more favorable.

The treatment of water molecules which hampered the prediction of the correct binding mode for compound **15** and also proved essential for correct prediction of the binding mode for compound **4**, is known to pose a serious challenge in molecular docking (de Beer et al., 2010). It is somewhat unsurprising that this problem was encountered since the effect of ion and water association with RNA molecules has been found to be of great importance for ligand binding (Fulle and Gohlke, 2010).

For the best scoring decoy, compound **6**, no such incompatibility of the binding site could be identified. Instead, the protonation state was not accounted for correctly, which led to incorrect predictions. This problem is well known for protein-ligand docking and is described in the literature (ten Brink and Exner, 2009). The conclusion to be drawn from this is that the rule-based generation of relevant protonation states could be improved. Reliable computational protocols for the prediction of protonation states and tautomeric states of small ligands are still under development and are not fully mature yet (Greenwood et al., 2010). Wherever possible experimental data on protonation states and tautomeric states should be included for the preparation of screening libraries.

For the assembly of the screening set of compounds the in-house database of

commercially available compounds was filtered for suitable compounds. The filter was designed to ensure that the compounds in the data set were chemically diverse while resembling the known binders in very fundamental properties. The careful selection of screening compounds and the removal of compounds with a positive van der Waals score after docking avoided the phenomenon of overenrichment (Verdonk et al., 2004). In case of over-enrichment the screening set includes a large number of compounds which cannot bind the target and which could be excluded by much simpler methods than docking. Following is a brief discussion of the individual filter categories used to assemble the screening set.

The limitation to a maximum of 18 non-hydrogen atoms ensures molecules do not exceed the binding site size. This is only a very rough estimate, and by no means ensures that all molecules fit into the binding site, since molecule shape is not taken into account. But it is expected that no molecules of more than 18 non-hydrogen atoms fit into the binding pocket in the conformation present in the crystal structure used for docking. The known ligands present in the test set (Tab. 3.2) did not exceed eleven non-hydrogen atoms.

Since the *in vivo* ligand adenine is bound in the binding pocket forming multiple hydrogen bonds with acceptor and donor sites on both the ligand and the riboswitch, it can be inferred that any ligand will at least have one hydrogen bond donor and acceptor site. Also the lack of interaction partners for too many functional groups within the binding site will result in very weak affinities. The near neutrality of the putative ligands is a requirement resulting from the neutrality of the binding site. The whole binding site does not have a net charge and no formal charges can be found on any of the functional groups interacting with the native ligand adenine. In order to explore the ability of the binding site to accommodate charged ligands, compounds bearing a single negative or up to a double positive charge were taken into account. Interestingly, none of the molecules that were considered for experimental testing was charged. The overall neutrality of the binding site seems to have selected strongly for neutral ligands.

The requirement of one or two ring systems was motivated by the observation that all known ligands for this riboswitch contain a ring system. This ring system in all known cases is aromatic and the ring is stacked in between nucleo-bases for all known crystal structures. Therefore, it was concluded that any potential ligand would have to contain a ring system.

When docking a large set of compounds into the GRA binding site, the known ligands accumulated in the top part of the score-ranked list (Fig. 3.4). This is the desired outcome for this experiment. Known binders should have a favorable score and rank high in a chemically diverse set of compounds where only a small minority will bind to the riboswitch target. For the decoys an enrichment is achieved as well. Ideally a set of decoys would not be enriched and an AUC of 0.5 would be obtained in a ROC plot. This would reflect a random distribution of the decoys in the score-ranked list. The enrichment of the decoy set in this situation has been attributed to the similarity of the decoys to known ligands. Most decoys only differ from ligands by a single functional group. In contrast the screening set in whole is much more chemically diverse, so the decoys used are not spanning the whole chemical space of the screening library. A slight enrichment of the decoys in these circumstances is unsurprising and unproblematic as long as discrimination of ligands and decoys is achieved. Still even though the decoys are enriched, they span the upper 50% of the score-ranked docking list almost evenly. In contrast the ligands only span the upper 20% and all but one are found in the top 2% of the screening set (Fig. 3.4). So, in summary separation of ligands and decoys is achieved and the performance of the docking protocol is regarded as satisfactory.

3.3.3 Prospective results

While the ability to reproduce known binding data is encouraging, the real test for any docking algorithm is whether it can predict novel ligands for a given target.

This is the practical purpose of virtual screening after all and the common application for docking in the drug discovery context. To test the predictive power of the docking algorithm used, five compounds from the 25 best-scored compounds were selected for experimental characterization after manual inspection of the predicted binding modes. From these five compounds three were closely resembling known ligands and two compounds based on scaffolds not found in known ligands (Tab. 3.3). Four of the selected five compounds were confirmed to bind the riboswitch, which is a very encouraging success rate.

The only selected compound, **25**, that was found not to bind the riboswitch closely resembled previously known ligands. Compound **25** is based on the same scaffold as the triazine ligands and very similar to the pyrimidine ligands. It differs from compound **2** only in the replacement of one extra cyclic amino group by a chlorine atom. None of the ligands in the test set contain a chlorine atom while two of the decoys do. Therefore this was considered a very interesting test case. There are two possible reasons for the observed lack of binding: the possible locations for the chlorine atom would be in close proximity to oxo-groups from surrounding bases, while the angular requirements for a C-Cl-O interaction (Bissantz et al., 2010) are not satisfied. The second factor is the inductive effect of the chlorine which is expected to act on the basicity of the ring nitrogens and thereby is expected to render them less basic. This results in weaker hydrogen bonds leading to a loss in binding affinity. The scoring function used here does not have the capability to include these effects in the predictions. While it would be beneficial to extend the scoring function to include angular effects, it is remarkable how well the scoring function performs without any terms considering angular dependencies. The inductive effect would have to be treated by a modified approach to ligand preparation, namely a more sophisticated protocol for the assignment of partial charges to the ligand atoms.

Compound **27** differs from compound **2** only by the absence of a single amino group. The binding of this compound was expected, the observed fourfold drop

in binding affinity and the associated decrease in binding energy is very well explained by absence of the interactions from the missing amino group. The crystal structure of the complex of the riboswitch with compound **27** revealed a double binding mode. One of the observed binding conformations was the one predicted by the docking algorithm. The other binding mode was initially not predicted. This was, however, due to the setup of the docking algorithm, which only predicts one binding mode per compound in the screening set. Since upon consideration of more than one binding mode with favorable score both binding modes were predicted, the docking protocol was in principle capable of reproducing the observed double binding mode. The occurrence of multiple binding conformations with near-optimal characteristics is a well-known phenomenon in docking and protocols accounting for the associated effects are being developed (Gorelik and Goldblum, 2008; Stjernschantz and Oostenbrink, 2010). Thus, this is another example of RNA-ligand docking suffering from the same problems already known from protein-ligand docking.

The binding of compound **26** was not unsurprising given the similarity to compounds **2**, **5** (Tab. 3.2) and **27** (Tab. 3.3). It differs from compound **2** only by the replacement of one amino group by a methyl group. The measured affinity for compound **26** is very similar to compound **27**. Therefore it can be concluded that the methyl group does not contribute any significant binding energy but is also not detrimental to binding. Compound **26** has better shape complementarity to the binding pocket than compound **27**, however the water molecule near U22 needs to be shifted in location to make room for the methyl group. The binding mode was not correctly predicted by the docking algorithm, which is explained by the fact that the water molecule was kept in the original place during docking. This presents the third example apart from compound **4** and compound **15** where the correct binding mode prediction is sensitive to the water molecule used for docking.

Upon docking with the observed location for the essential water molecule, the

correct binding mode was predicted. This underpins the confidence in the scoring function but again points out a limitation of the used docking approach already observed for the binding of adenine. The location of water molecules and the prediction of their involvement in binding has proven to be a challenge for docking for a long time (see above).

Compound **28** was found to bind the riboswitch with the weakest affinity of the newly discovered ligands. Unfortunately all crystals that could be obtained were insufficient in size for data collection. Attempts to improve the crystals were unsuccessful and this was attributed to the solubility limit of the ligand. The molecular scaffold of compound **28** was not observed previously in GRA ligands. This ligand is the smallest known to bind this riboswitch and is based on a 1,2,4-thiadiazole moiety. Additionally, the two heterocyclic nitrogens and the exo-cyclic amino group closely resemble the guanidino-like moiety found in several known ligands including compounds **1**, **2**, **5** and **27**. Therefore it would have been interesting to study the adaptation of the binding pocket to accommodate this small ligand. The degree of flexibility for this binding pocket is yet to be determined, even though experimental evidence suggests the ability to adjust to various requirements for ligand binding (Edwards and Batey, 2009; Gilbert et al., 2009).

Compound **24** is the only ligand based on a novel scaffold for which a co-crystal structure could be determined. The binding affinity of 350 μM is rather weak but binding still could be confirmed. The binding mode was predicted correctly which boosts the confidence in the selected docking protocol. When comparing the binding mode of compound **24** to the riboswitch with the binding mode of the original ligand from the crystal structure a great similarity in the interaction pattern becomes apparent (Fig. 3.9). This is somewhat surprising since the chemical structure of the two compounds differs greatly. While compound **24** is based on two fused five-membered rings the original ligand from the crystal structure is based on a pyrimidine scaffold. Both ligands interact with the Watson-Crick

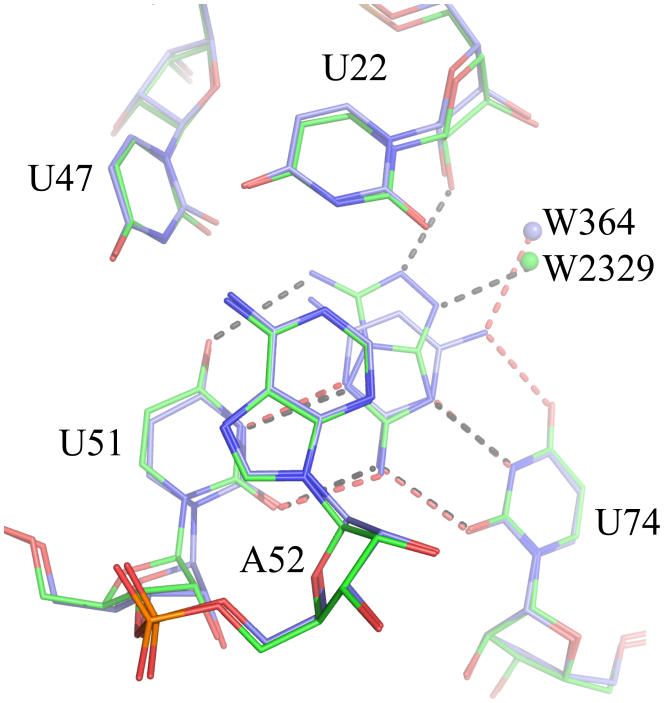


Figure 3.9: Comparison of the binding modes of compound **5** (RNA and ligand in blue carbon atoms, water molecule in blue, hydrogen bonds in red) and compound **24** (RNA and ligands in green carbon atoms, water molecule in green, hydrogen bonds in black). Overall the ligand-receptor interactions are very similar for both ligands. The hydrogen bonds between compound **5** and U51 are parallel and have near optimal geometry while for compound **24** the hydrogen bonds to U51 are not parallel and have less optimal geometry. Also compound **5** has an amino group forming a hydrogen bond with the U74 O4 while compound **24** does not form a hydrogen bond with this group.

edges of U51 and U74 and the water molecule 364 in case of triaminopyrimidine or water molecule 2013 in case of compound **24**, respectively. U51 forms three hydrogen bonds with both ligands but in case of compound **24** these are slightly deviating from the ideal angle of a hydrogen bond (Fig. 3.9). The distance of the two exo-cyclic amino groups is greater than required for optimal hydrogen bonding (Fig. 3.9). U74 forms three hydrogen bonds with triaminopyrimidine while only two are formed with compound **24**. This lack of one hydrogen bond and the less ideal hydrogen bonds to U51 are likely the reasons for the reduced affinity of compound **24** compared to compound **5**. Both ligands interact with the water molecule close to U22. Interestingly, the location of the ligand atom

forming the hydrogen bond varies between the ligands while the water molecule is in a very similar location in both cases. Compound **24** forms one extra hydrogen bond to the riboswitch compared to compound **5**. This hydrogen bond is formed between the 2'-hydroxyl of U22 and the N2 of the ligand.

Overall, the prediction of ligand binding modes proved to be accurate provided the relevant binding site conformation is used for docking. Compound **24** was predicted correctly right away and the observed binding modes for compounds **26** and **27** were predicted correctly after providing the correct location of the crucial water molecule and allowing for multiple binding modes. In future development these issues will need to be addressed to ensure accurate predictions in a wider set of circumstances. The ability to correctly predict the binding modes of novel ligands is the most important feature of any docking protocol. In a drug discovery context virtual screening by molecular docking is used to identify novel ligands. Even successful prediction of ligand binding, but with an incorrect binding mode, is of limited value. The agreement of predicted and observed binding mode is an indication that the prediction is real and not just a coincidence. It has previously been observed that even though binding is correctly predicted for a compound to a given target, the observed binding mode significantly deviates from the predicted one (Mpamhanga et al., 2009). In such cases the docking still needs to be considered a failure, since the prediction was based on incorrect calculations.

3.3.4 Conclusions

In the reported approach on RNA-ligand docking known experimental data was successfully reproduced and new ligands were predicted successfully. The reproduction of known binding data was a first success which was underpinned by the successful identification of several causes for mispredictions. The rationalization of the mispredictions and the successful predictions of the real binding modes upon correction point to a good choice of the GRA as a model system.

Identification of novel ligands proved the usefulness of the protocol in a drug discovery setting. However, not unexpectedly several problems with RNA-ligands were identified despite of the overall success. Surprisingly, none of these problems were RNA-specific but the identified problems are well known from protein-ligand docking. The treatment of water molecules has been a substantial challenge and is still a current topic for research. Also the treatment of tautomeric states and protonation states has been previously known to be an issue (ten Brink and Exner, 2009; Greenwood et al., 2010). Theoretically, the energy of all possible states would have to be determined and each binding score would have to be adjusted by the energy of that state. In practice the tautomers and protonation states relevant to docking are often determined by a rule-based approach as in this study.

Finally, it is important to note that the above discussion is referring to the binding properties of the isolated aptamer domain only. Activity on the full riboswitch including the expression platform *in vitro* and *in vivo* of these compounds is yet to be tested. Especially for riboswitches under kinetic control, the binding affinity could deviate significantly from the concentration required to alter gene expression.

Table 3.1: Statistics of data collection and refinement for the crystal structures of GRA-ligand complexes

Ligand Complex	**3**	**15**	**24**	**26**	**27**
PDB code	N/A	2xo1	2xnw	2xnz	2xo0
Space group	C2	C2	C2	C2	C2
Unit cell dimensions	a=131.910	a=132.762	a=135.895	a=132.165	a=130.923
	b=35.110	b=35.180	b=35.380	b=35.068	b=34.940
	c=41.800	c=41.763	c=42.197	c=41.778	c=42.074
	β=90.10	β=90.51	β=92.20	β=92.06	β=92.11
Resolution range[a] (Å)	60-1.8 (1.88-1.80)	20-1.6 (1.66-1.6)	67.88-1.5 (1.55-1.5)	60.08-1.6 (1.66-1.6)	65.5-1.7 (1.76-1.7)
Observations	49,180	72,587	80,100	93,460	62,872
Unique observations	17,250	25,098	30,250	25,334	21,077
Completeness (%)	93.0 (71.1)	96.5 (91.0)	93.1 (97.9)	98.5 (98.1)	96.8 (93.5)
$\left\langle \frac{I}{\sigma(I)} \right\rangle$	27.9 (4.1)	20.6 (2.7)	18.0 (2.3)	30.2 (2.9)	26.8 (3.1)
R_{merge}[b] (%)	6.4 (34.9)	5.3 (25.8)	6.2 (38.5)	4.4 (39.8)	5.2 (39.1)
Refinement statistics					
Resolution range[c] (Å)	60-1.8	20-1.6	67.88-1.5	60.08-1.6	65.5-1.7
R-factor[d] % (R_{work}/R_{free})	18.45/22.2	19.5/21.1	21.4/24.4	19.8/23.4	21.6/27.1
Number of atoms[e]	1386/10 228/60	1364/11 226/56	1366/10 329/54	1382/9 216/52	1363/16 151/55
Mean B-factor (Å^2)[f]	17.8/24.4 23.3/30.5	22.7/18.3 29.0/25.8	21.7/22.1 29.0/23.8	24.9/22.7 29.2/27.9	29.6/26.3 30.2/32.7
RMS bond length deviation (Å)	0.014	0.008	0.009	0.009	0.013
RMS bond angle deviation (°)	2.135	1.508	1.683	1.546	1.973

[a]Values in brackets are for the highest resolution shell.
[b]$R_{merge} = \sum (|I - \langle I \rangle| / I)$
[c]Values in brackets are for the highest resolution shell.
[d]R-factor=$\sum |F_o - F_c| / \sum F_o$
[e]Number of atoms for RNA, ligand, water molecules and other atoms, respectively
[f]Mean B-factors for RNA, ligand, water molecules and other atoms, respectively

Table 3.2: Test set of experimentally confirmed ligands and decoys taken from the literature. All K_D values quoted are from the respective publications (Mandal and Breaker, 2004; Gilbert et al., 2006a).

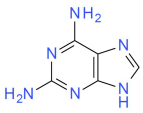
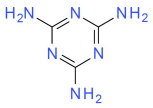
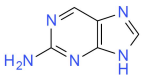
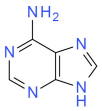
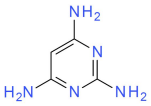
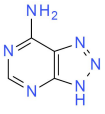
Compound No.	Structure	Rank	Score $[\frac{kJ}{mol}]$	RMSD [Å]	$K_D[\mu M]$
1		1	-34.18	0.24	0.01
2		2	-33.97	NA	20
3		3	-32	NA	0.3
4		4	-31.85	0.34	0.3
5		5	-31.81	0.13	20
6		6	-30.44	NA	NA
7		7	-29.90	NA	30

Table 3.2: Test set of experimentally confirmed ligands and decoys taken from the literature. All K_D values quoted are from the respective publications (Mandal and Breaker, 2004; Gilbert et al., 2006a) (continued).

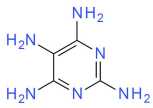
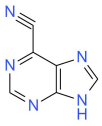
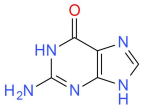
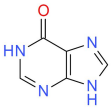
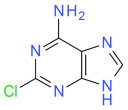
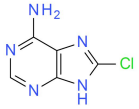
Compound No.	Structure	Rank	Score $[\frac{kJ}{mol}]$	RMSD [Å]	$K_D[\mu M]$
8		8	-26.71	NA	20
9		9	-25.11	NA	NA
10		10	-23.24	NA	NA
11		11	-21.36	NA	NA
12		12	-20.03	NA	NA
13		13	-18.99	NA	NA
14		14	-18.33	NA	NA

Table 3.2: Test set of experimentally confirmed ligands and decoys taken from the literature. All K_D values quoted are from the respective publications (Mandal and Breaker, 2004; Gilbert et al., 2006a) (continued).

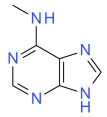
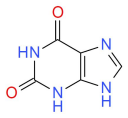
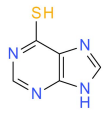
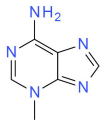
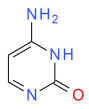
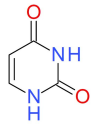
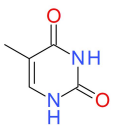
Compound No.	Structure	Rank	Score $[\frac{kJ}{mol}]$	RMSD [Å]	$K_D[\mu M]$
15		15	-17.49	NA	100
16		16	-15.97	NA	NA
17		17	-14.48	NA	NA
18		18	-14.17	NA	NA
19		19	-11.97	NA	NA
20		20	-9.93	NA	NA
21		21	-8.50	NA	NA

Table 3.2: Test set of experimentally confirmed ligands and decoys taken from the literature. All K_D values quoted are from the respective publications (Mandal and Breaker, 2004; Gilbert et al., 2006a) (continued).

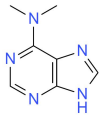
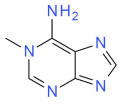
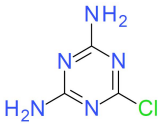
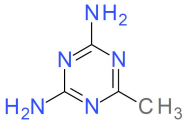
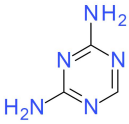
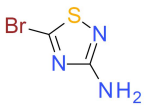
Compound No.	Structure	Rank	Score $[\frac{kJ}{mol}]$	RMSD [Å]	$K_D[\mu M]$
22		22	-7.01	NA	NA
23		23	15.92	NA	NA

Table 3.3: Compounds selected and subsequently tested for binding and subjected to crystallography.

Compound No.	Structure	Rank	RMSD [Å]	$K_D[\mu M]$
24		3	0.21	340 ± 40
25		7	NA	NA
26		11	0.29	110 ± 30
27		14	0.65/0.23	80 ± 20
28		33	NA	650 ± 180

RMSD in Å between docking prediction and binding mode from the crystal structure

Chapter 4

Virtual screening for the *T. tengcongensis* SAM-I riboswitch

4.1 Overview

To apply RNA-ligand docking in a drug discovery scenario, virtual screening for novel ligands for the SAM-I riboswitch was performed. First, the docking protocol was validated for this target by reproduction of crystallographically determined binding modes. Then promising potential binders were identified from the docking results of a large set of compounds. Compounds were then purchased and tested for binding. One novel binder could be identified among the tested ones.

4.2 Results

4.2.1 The SAM binding site

The SAM-I riboswitch was selected for this virtual screening study since it was implicated to be a potential drug target (Blount and Breaker, 2006). The sequence of the *T. tengcongensis* SAM-I riboswitch is similar to the SAM-I riboswitches in

pathogenic bacteria. Also the binding site did offer challenges beyond the model binding site of the GRA (Daldrop et al., 2011). The SAM-I binding pocket is considerably larger, binds to flexible ligands and is partly solvent accessible (Fig. 1.8; Montange and Batey 2006). Additionally, binding data and crystal structures for the SAM-I riboswitch and several ligands were available to validate the initial docking setup. This also implied that the SAM-I riboswitch would be amenable to cocrystallization with potential novel ligands.

4.2.2 Docking of SAM analogues

To validate the docking setup, ligands for which the binding modes are known are usually redocked into the receptor binding site (Hartshorn et al., 2007). In case of this riboswitch docking of the natural SAM ligand (Fig. 4.1) into the SAM-I riboswitch binding pocket was not possible. The protocol for generation of a dockable database does not allow for a positively charged trivalent sulfur atom to be present in the ligand because the force fields used for ligand preparation contain no parameters for this atom. The conformational flexibility of the compounds to be docked was taken into account by docking an ensemble of precalculated low-energy conformers, which were generated and scored with the MMFF94 force field (Halgren, 1996). This force field does not contain any parameters for the sulfur atom contained in the SAM-ligand. However, analogues of SAM compatible with the docking protocol had been crystallized in complex with the SAM-I riboswitch (Montange et al., 2010). These were used to validate the receptor setup and the docking protocol. These analogues are S-adenosylhomocystein (SAH) and sinefungin (SFG), also known as S-adenosylornithine (Fig. 4.1). The binding modes of the SAM analogues SAH and SFG are almost identical to the binding mode of SAM (Fig. 4.2) which is reflected in an RMSD of 0.5 Å (based on all non-hydrogen atoms present in both structures) between SAM and SAH, and an RMSD of 0.45 Å between SAM and SFG. SAH differs from SAM only by the

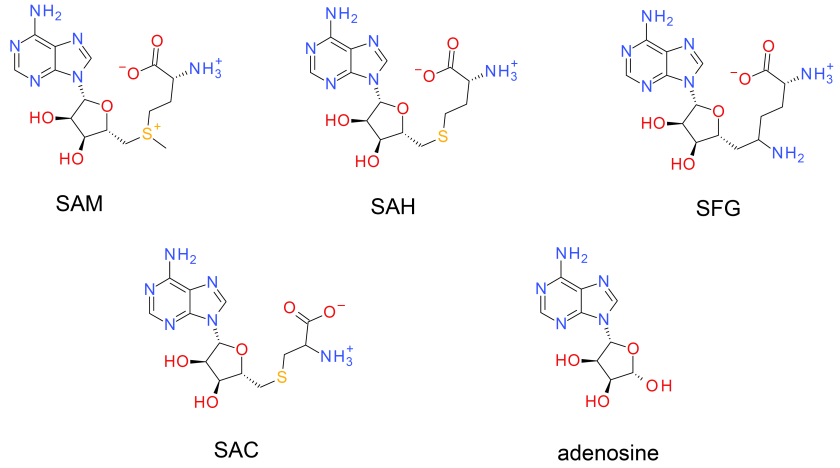


Figure 4.1: Chemical structure of SAM and SAM analogues used to validate the docking setup by redocking into the SAM-I binding site.

absence of the methyl group attached to the sulfur atom. This relatively small change has one important consequence: the sulfur atom now is only connected to two neighboring atoms and is uncharged. The affinity of SAH for the SAM-I riboswitch was reported as approximately 100 fold weaker than the affinity for SAM (Montange et al., 2010; Winkler et al., 2003). In SFG the SAM-sulfur is replaced by a carbon atom and the methyl group is replaced by an amino group (Fig. 4.1). The binding constant was reported to be 180-fold weaker than for SAM (Montange et al., 2010). To these two analogues S-adenosylcystein (SAC) was added for which no structural data was available. However, for the *B. subtilis yitJ* SAM-I riboswitch SAC was reported to bind with 30 μM affinity (Winkler et al., 2003). Due to its highly similar chemical structure SAC was assumed to exhibit a binding mode closely resembling that of SAH.

Docking of these ligands was performed using the same protocol as described for the GRA (Chapter 3). The sphere set to define the binding pocket was limited to the area usually occupied by the SAM-adenine. The docking results for SAC, SAH and SFG vary in their agreement with the known or expected binding modes. Docking of SFG did not reproduce the crystallographically observed binding mode (Montange et al., 2010). Not even substructures of the ligand were placed cor-

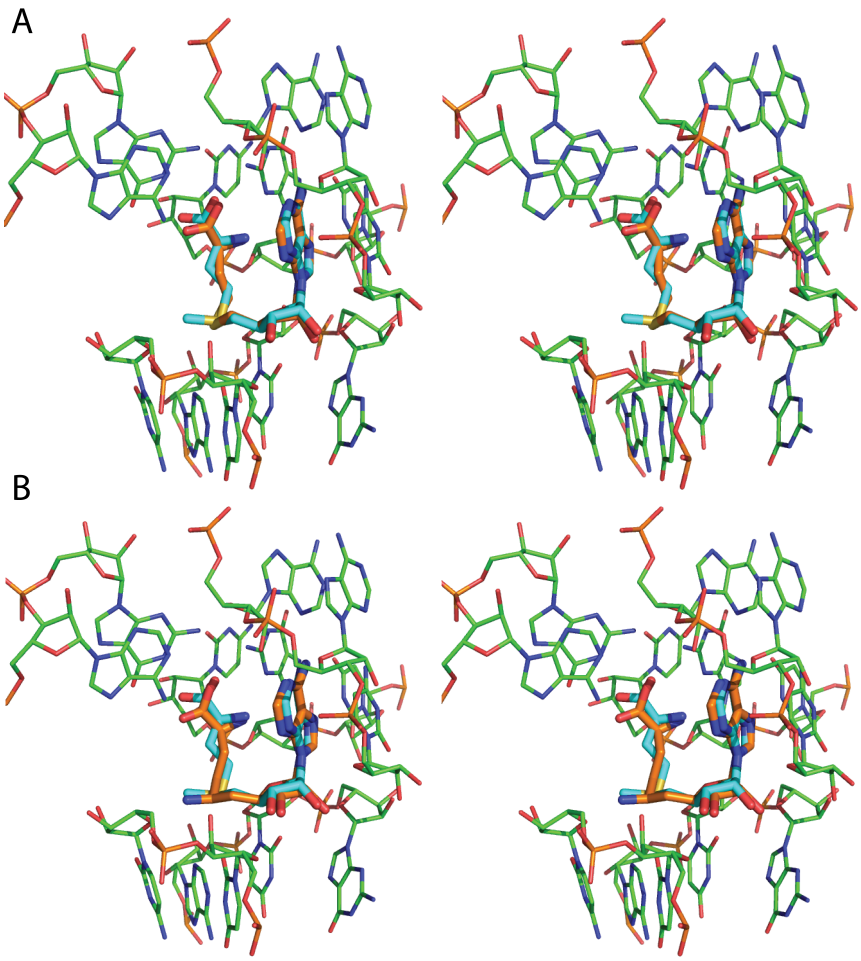


Figure 4.2: Comparison of the binding mode for SAM and SAM-analogues to the SAM-I riboswitch. A: Wall-eyed stereo view of the superposition of the binding mode of SAH (amber carbon atoms) and SAM (amber carbon atoms) in the SAM-I binding site (green carbon atoms) from PDB-file 2gis. B:Wall-eyed stereo view of the superposition of the binding mode of SFG (amber carbon atoms) and SAM (amber carbon atoms) in the SAM-I binding site (green carbon atoms) from PDB-file 2gis.

rectly (not shown). For SAH, the predicted binding mode agreed with the observed one at least with respect to the general location of the adenine part of the ligand (Fig. 4.3).

For SAC, for which no structural data was available, the predicted binding mode was very similar to the one expected based on the observed binding mode for SAH. Particularly the adenosine moieties aligned well with an RMSD of <0.8 Å. Considering these results one trend became apparent: The agreement between

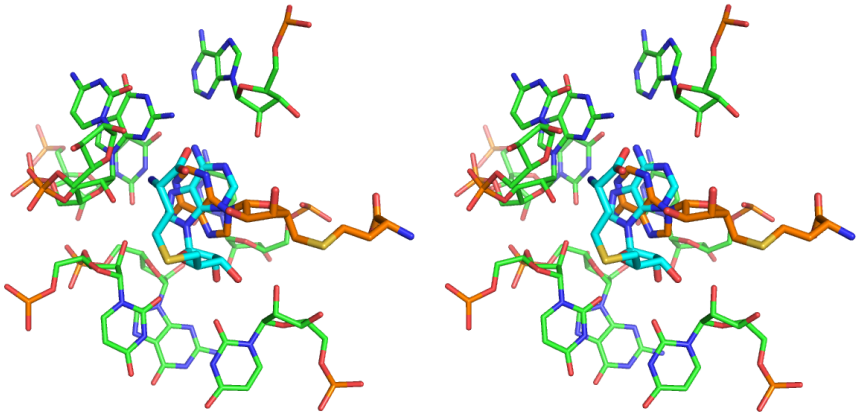


Figure 4.3: Wall-eyed stereo view of the predicted binding mode of SAH (amber carbon atoms) in the SAM-I riboswitch binding pocket (green carbon atoms). The experimentally observed binding mode for SAH is shown in cyan carbon atoms.

experimental data and prediction is best for the SAC which contains the shortest amino acid moiety and deteriorated with increasing size of the amino acid moiety of the ligand. As a close inspection of the conformational ensembles generated for these ligands revealed, this was caused by insufficient conformational sampling. Two measures were taken to validate the docking setup despite this problem:

1. The ligand conformations for SAH and SFG observed in the crystal structure of the riboswitch ligand complex were docked into the binding pocket.

2. Adenosine (Fig. 4.1) was docked into the binding pocket.

Docking of the crystallographically observed conformation of SAM-I bound SAH and SFG resulted in predictions that were almost identical to the known binding modes (Fig. 4.5). The RMSD between prediction and known binding mode was 0.24 Å for the SFG and 0.23 Å for the SAH. The docking scores of these conformations were more favorable than the ones obtained in the previous docking.
Next, adenosine was docked into the SAM-I riboswitch. The predicted binding mode was in agreement with the the adenosine parts of the known ligands (SAM, SFG and SAH). This is reflected in an RMSD of 0.25Å between the predicted binding mode for adenosine and the SAM-adenosine moiety from the crystal

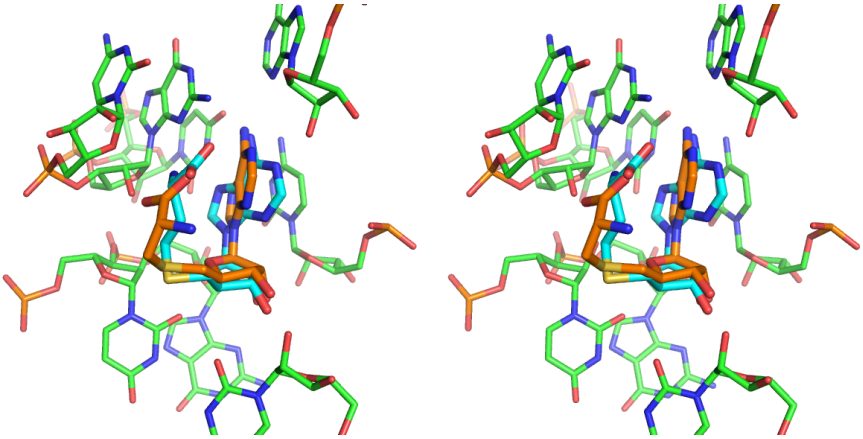


Figure 4.4: Wall-eyed stereo view of the predicted binding mode of SAC (amber carbon atoms) in the SAM-I riboswitch binding pocket (green carbon atoms). The experimentally observed binding mode for SAH as a comparison is shown in cyan carbon atoms. The adenosine parts of the experimentally determined SAH binding mode agrees well with the SAC binding mode prediction.

structure.

Taken together these results led to the conclusion that the failures in redocking were due to insufficient conformational sampling rather than the inability to correctly judge the ligand-receptor interactions. For the planned virtual screening this was not considered to be a relevant problem due to the nature of the compounds in the screening set (see below).

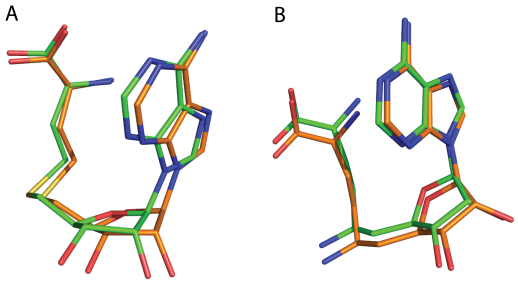


Figure 4.5: A: Superposition of the predicted binding mode for SAH using only the conformation observed in the crystal structure (amber carbon atoms) and the known binding mode(green carbon atoms). B: Superposition of the predicted binding mode for SFG using only the conformation observed in the crystal structure (amber carbon atoms) and the known binding mode(green carbon atoms).

4.2.3 Assembly of a screening set of compounds

A set of compounds to be docked into the SAM-I binding site was assembled in order to proceed with the virtual ligand screening. The in-house database of commercially available chemical compounds contained approximately 6.5 million unique chemical compounds (Brenk et al., 2008). Compounds were selected from this database using the following criteria:

- maximum of 26 non-hydrogen atoms
- maximum molecular weight of 500 Da
- minimum of 2 hydrogen bond acceptors and 2 hydrogen bond donors
- one or two ring systems
- net charge between 0 and +2
- maximum of 4 rotatable bonds
- no unwanted groups

The rationale for these criteria are as follows: A maximum of 26 non-hydrogen atoms and 500 Da molecular weight puts an upper limit on the size of the potential ligands. In drug discovery this is very common practice to ensure drug likeness (Lipinski, 2001). Next, the minimum of four functional groups capable of hydrogen bonding ensures that the compounds have the potential for multiple strong interactions with the SAM-I binding site which is required for a high affinity ligand. The minimum of one ring system is a requirement of the binding pocket since all known ligands contain an adenine moiety, and the area of the binding pocket where this moiety binds is prone to be occupied by a ring system. The requirement for the net charge was derived from the cationic nature of the natural SAM ligand which bears a net charge of +1. Since neutral ligands like SAH were reported in the literature (Montange et al., 2010; Lim et al., 2006), a

margin of ±1 in net charge was tolerated. A maximum of four rotatable bonds ensured that the conformational space could be sufficiently sampled, avoiding the complications experienced with the SAM-analogues. Finally, exclusion of compounds containing functional groups defined as unwanted ensured that the obtained compounds did not contain structural features which are known to be problematic in drug discovery (Brenk et al., 2008). In total, $\approx 95,000$ compounds passed all filters and constituted the screening set.

4.2.4 Analysis of docking results and selection of compounds for experimental characterization

The screening set was docked into the SAM-I riboswitch binding site. For docking the same sphere set was used that was also used for docking of the SAM analogues, only covering the area where the adenine moieties of the known ligands are located. A total of $\approx$ 66000 compounds was successfully placed into the binding pocket and scored. The score-ranked list of docked compounds was analyzed to identify candidates for experimental study. To guide the selection of compounds for experimental testing, the list was further filtered using a pharmacophore (Fig. 4.6). The pharmacophore was developed based on a three hydrogen

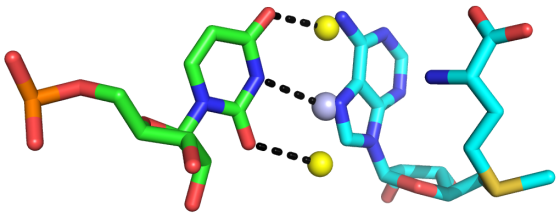
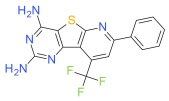
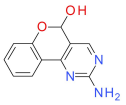
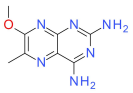
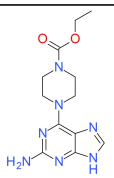
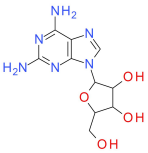
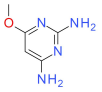
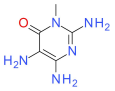
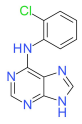


Figure 4.6: Pharmacophore for filtering of the SAM-I docking results. The crucial nucleotide U57 is shown in green carbon atoms, the SAM-ligand in cyan carbon atoms. The pharmacophore is composed of two donor groups (yellow spheres) and an acceptor functionality (grey sphere).

bond interaction pattern with U57. Donor functionalities were placed close to the O2 and O4 of U57 and an acceptor functionality close to the N3 of U57 (Fig. 4.6). The docking results were subsequently filtered requiring two of these three molecular recognition characteristics to be present. A total of 307 compounds matched these requirements.

Table 4.1: Compounds selected from the docking results for the SAM-I riboswitch and subsequently tested for binding.

Compound No.	Structure	Score	Compound No.	Structure	Score
1466736		-36.58	2071368		-32.16
1672502		-30.90	4051493		-34.03
1820323		-30.33	5155566		-22.44
1946503		-22.82	5669778		-32.93

Finally a set of eight compounds was selected based on visual inspection of the 307 binding mode predictions for experimental testing and ordered from the respective suppliers (Tab. 4.1). The selected compounds differed in molecular weight ranging from 140 Da for the smallest compound to about 365 Da for the

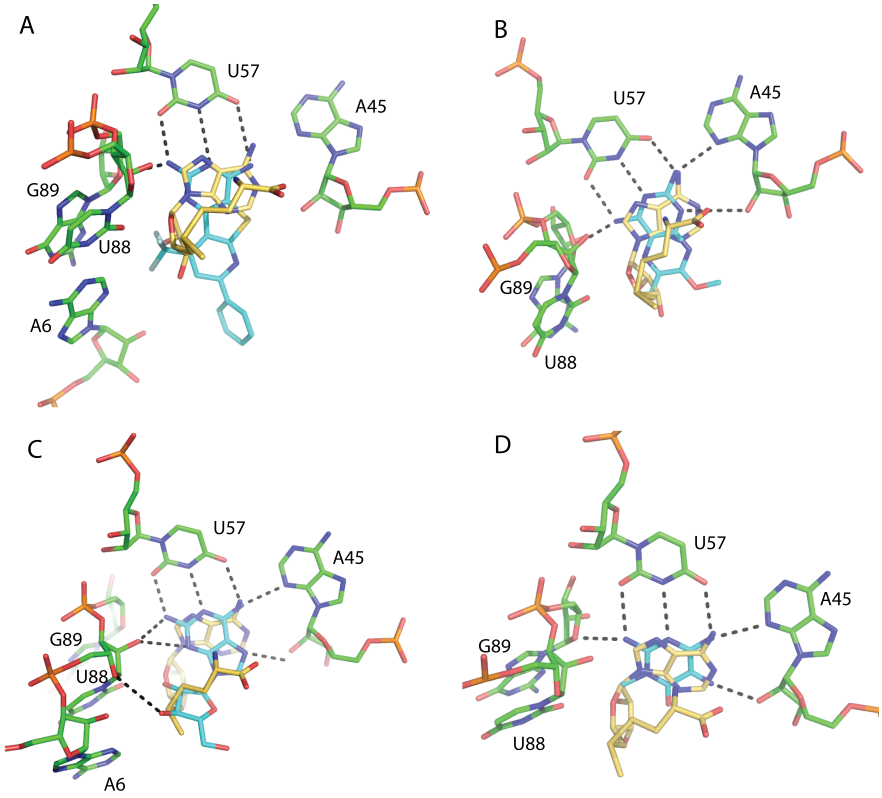


Figure 4.7: Binding mode predictions for selected compounds (Tab. 4.1) from a virtual screening of the SAM-I riboswitch. The key binding site nucleotides from the riboswitch structure (PDB-file 2gis) are shown in green carbon atoms. The binding mode predictions are shown in cyan carbon atoms with the predicted hydrogen bonds between compound and receptor indicated as black dashes. The SAM-ligand is shown with yellow carbon atoms. A: Prediction for compound 1466736, B: Prediction for compound 1672502, C: Prediction for compound 1820323, D: Prediction for compound 1946503.

largest compound and were based on a variety of molecular scaffolds. The predicted binding modes for the selected compounds are shown in figures 4.7 and 4.8. As required by the pharmacophore all compounds form multiple hydrogen bonds with U57 and a heterocycle occupies the area in which the adenine moiety of the SAM is bound in its complex structure. From this selection compound 1820323 (Tab. 4.1 and Fig. 4.7) requires a closer inspection. This compound differs from adenosine (Fig. 4.1) only in two ways, first the nucleobase carries an additional amino group at the C2 position and second the sugar moiety differs in its stereo chemistry. In the docking prediction the sugar is an arabinose rather than the

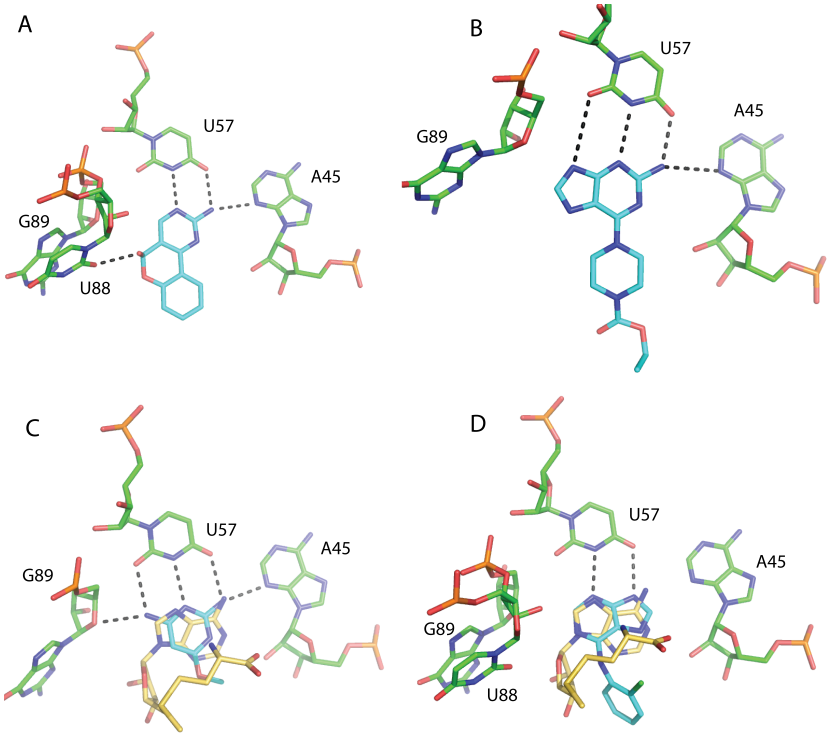


Figure 4.8: Binding mode predictions for selected compounds (Tab. 4.1) from a virtual screening of the SAM-I riboswitch. The key binding site nucleotides from the riboswitch structure (PDB-file 2gis) are shown in green carbon atoms. The binding mode predictions are shown in cyan carbon atoms with the predicted hydrogen bonds between compound and receptor indicated as black dashes. Where appropriate the SAM-ligand is shown with yellow carbon atoms. A: Prediction for compound 2071368, B: Prediction for compound 4051493, C: Prediction for compound 5155566, D: Prediction for compound 5669778.

ribose observed in standard nucleic acids. Unfortunately, the arabinose variant of this ligand was not available commercially, so the ribose variant of this compound (also known as 2-aminoadenosine) differing only in the stereo conformation of the C2' was bought for experimental testing.

The ribose variant of compound 1820323 was also docked into the SAM-I binding site and obtaining a score of -26.39 $\frac{kJ}{mol}$ which is less favorable than the -30.33 $\frac{kJ}{mol}$ for the arabinose version. The predicted binding mode is shown in figure 4.9.

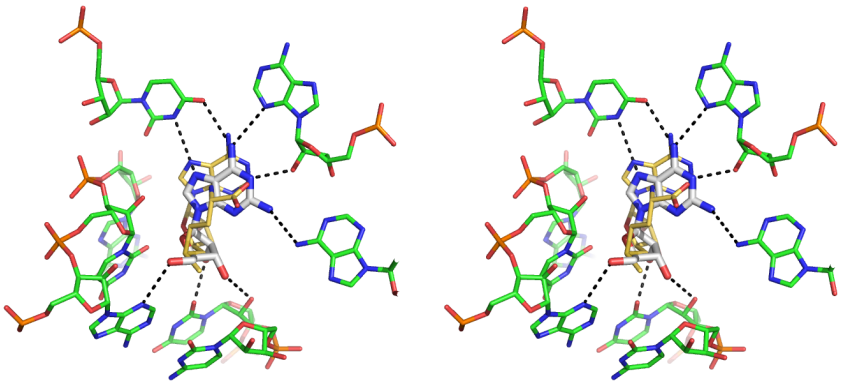


Figure 4.9: Wall-eyed stereo view of the predicted binding mode of 2-aminoadenosine (grey carbon atoms) to the SAM-I riboswitch (green carbon atoms) with the hydrogen bond interactions highlighted as black dashed lines. The SAM ligand from the crystal structure is shown in yellow carbon atoms.

4.2.5 Experimental testing of the selected compounds

The selected compounds were then tested for SAM-I binding using ITC (Wiseman et al., 1989). Since the affinity for these compounds was expected to be very low, a direct ITC approach was considered inadequate. Therefore, the ITC measurements were carried out as a competition experiment (Zhang and Zhang, 1998; Turnbull, 2005), where SAM was titrated into a solution of SAM-I riboswitch in absence and presence of the compound to be tested. In case of the compound binding the SAM-I riboswitch, the obtained apparent affinity for SAM to the riboswitch in the presence of that compound would be lowered. For the ITC experiments compounds had so be solubilised in DMSO at a concentration of 100 mM. Unfortunately only four of the eight compounds selected were soluble at this concentration and could be tested for binding by ITC. The compounds tested were:

- Compound 1820323 (Fig. 4.9)
- Compound 4051493 (Fig. 4.8 B)
- Compound 5155566 (Fig. 4.8 C)
- Compound 5669778 (Fig. 4.8 D)

From the four compounds tested only compound 5669778 was observed to bind to the SAM-I riboswitch. The titration data recorded in the absence and presence of compound 5669778 are shown in figure 4.10. The apparent affinity obtained from the fit of the data was 120 nM in the absence and 740 nM in the presence of 2 mM compound 5669778. The binding affinity for this compound was calculated to be 380 μM.

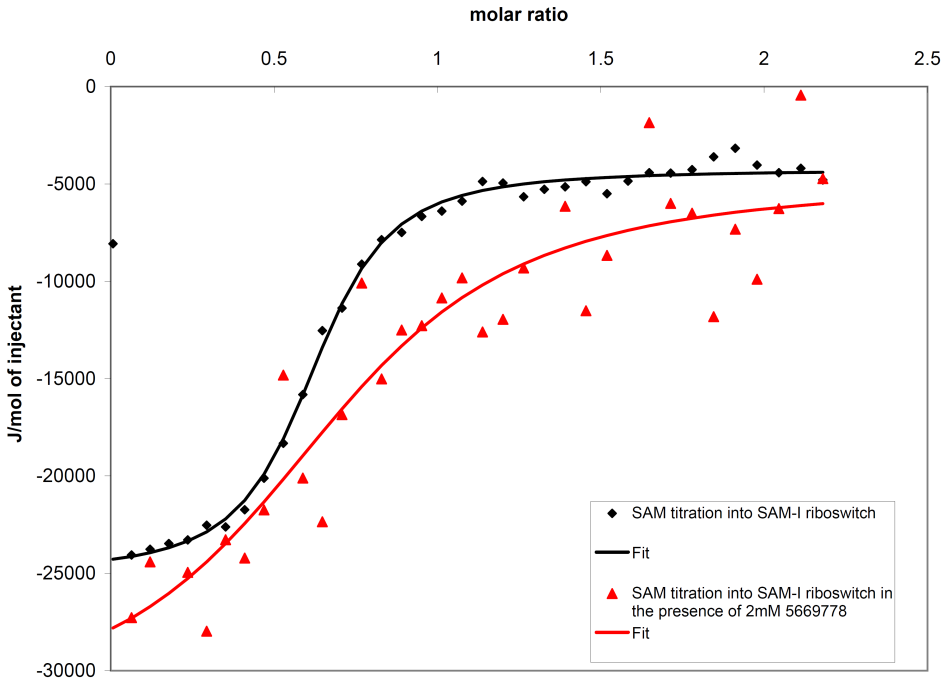


Figure 4.10: Isothermal titration calorimetry of SAM binding to the SAM-I riboswitch in competition with compound 5669778. SAM is titrated to a solution of SAM-I riboswitch in the absence and presence of 2 mM compound 5669778. Data were then fitted to a two state binding model and the binding constant of compound 5669778 was calculated from the apparent decrease of SAM affinity in the presence of compound 5669778. The so calculated binding constant was 380 μM.

In parallel cocrystallisation of the SAM-I riboswitch with each of the eight selected compounds was attempted. Crystals were obtained in the presence of all compounds except for compounds 1672502 and 1946503. However, crystal structure analysis revealed that the crystals only contained the apo-structure (Fig. 1.9)

4.3 Discussion

4.3.1 Docking setup

The SAM-I riboswitch was chosen for this docking study since it was expected to be a more challenging target than the purine riboswitch. Moreover, the SAM-I riboswitch was implicated as a potential antibacterial target and was found multiple times in the genome of bacteria, especially firmicutes. Identification of novel binders would be a first step in exploitation of this riboswitch as an antimicrobial target and offer potential applications.

Unfortunately the protocol used to prepare the putative ligands did not allow for direct docking of the *in vivo* ligand SAM. The sulfonium group which was prohibitive for docking is a rare occurrence in libraries for virtual screening. Our in-house database of commercially available compounds contains only 245 examples in about 6.5 million compounds. Compounds of this type would usually be considered undesirable in drug discovery because of their instability. While in theory the protocol could be adapted to allow this type of compounds the required effort outweighs the benefit by far.

Docking of the SAM-analogues was hampered by the insufficient sampling of conformational space for SAH and SFG. The required conformations were absent in the precalculated ensemble of conformations in the dockable database. The ability of the protocol to correctly predict the binding mode of SAH and SFG in principle was demonstrated by redocking of the conformations observed in the crystal structure. Since these predictions also received more favorable scores than the predictions from the initial docking run, the problem can be pinpointed to the generation of a complete ensemble of conformations. Further evidence for this interpretation stems from the prediction of the binding mode for adenosine. Adenosine is far less flexible than SAH or SFG and the predicted binding mode aligns well with the adenosine moieties of SAM, SAH and SFG. This suggests that RNA-ligand interactions are judged correctly. When docking flexible ligands

special care should be taken in the future to avoid this problem. It is noteworthy that the insufficient sampling of conformational space for the SAM-analogues is in no way a specific problem of RNA-ligand docking (Agrafiotis et al., 2007). In fact, the problematic part of the protocol is completely independent of the receptor. Nevertheless a more suitable treatment of highly flexible ligands would be desirable to solve the problem.

4.3.2 Docking of a large set of compounds

The screening set was assembled by filtering the available commercial compounds for desirable criteria and finally contained just under 100,000 compounds. After filtering the obtained docking results with a pharmacophor, 307 unique ligands remained, which were visually inspected.

The eight compounds selected all were restricted to the area of the binding pocket usually binding the adenosine part of the SAM-ligand. None of the ligands spanned the whole area usually occupied by the SAM. This is to a large degree a consequence of the used sphere set, which was restricted to the adenine area, in combination with the pharmacophore filtering which enforces two hydrogen bonds between the compounds and U57. In principle, interactions with nucleotides outside this area are not selected against, but would be considered desirable. One way of addressing this issue would be assembly of a second screening set with relaxed criteria for the compound filter to include compounds which would be able to interact with these different parts of the binding pocket. Especially allowing for a maximum of approximately 30 non-hydrogen atoms and relaxing the restriction of a maximum of two ring systems appear valid modifications of the filter.

4.3.3 Experimental binding data

In the course of this study one novel ligand for the SAM-I riboswitch was discovered. Form the eight compounds selected only four were amenable to ITC measurements due to issues with solubility. Poor solubility of compounds is a well known problem in drug discovery (Lipinski, 2001) and there is no indication that the solubility issues experienced in this study are not within expectations. To overcome these solubility issues, the following approaches might be used. First, the ITC assay itself still has some room for improvement, using higher RNA concentrations would allow for lower compound concentrations by improving the signal to noise ratio. This would allow detection of smaller differences in apparent SAM affinity, which in turn would allow for a lower concentration of the competing compound. However, since already 10 nmol of riboswitch RNA per titration are used, increasing the RNA concentration will make the assay very expensive and time consuming. To overcome this limitation a newer generation of ITC instruments, such as the ITC200 instrument (GE Healthcare) could be used which requires approximately seven-fold less sample.

Second, different solvents could be explored, which might solve the problem by allowing solubilisation of compounds not soluble in DMSO. Third, different binding assays could be used, not requiring such high ligand concentrations. Currently no obvious alternative stands out but possible options include NMR-based assays or an in-line probing assay.

From the compounds tested which did not bind the riboswitch, compound 1820323 (Tab. 4.1) is closely related to adenosine and is structurally related to the SAM ligand. However the result from the screen was based on an arabinose sugar, while the only commercially available similar compound which was then tested was based on a ribose sugar. The compound was purchased and tested despite the difference in stereo chemistry because of the high structural similarity to the adenosine substructure in the SAM ligand. Unfortunately no binding

data is reported on adenosine but only on AMP, which is negatively charged. The reported lack of binding for AMP would be expected while adenosine might still bind with weak affinity. To obtain a more complete picture whether the SAM-I riboswitch is capable of binding ribosides, adenosine binding needs to be investigated.

Compound 5155566 (Tab. 4.1) was the smallest of the ligands selected from the screen. It also received the least favorable score in the set of compounds selected for experimental characterization. This was however believed to be compensated by its small size, according to the principle of ligand efficiency (Hopkins et al., 2004). It would have been interesting to see if and how the binding site adapts to accommodate for such a small ligand. But since only a small part of the binding pocket interacts with this ligand, it seems that the predicted interactions are not sufficient to stabilize the predicted ligand-riboswitch complex. From the apo-structure it is known that the binding site conformation in the bound and unbound states differs significantly (Fig. 1.9; Stoddard et al. 2010). The A46 adenine moiety needs to be displaced from the binding pocket to allow binding of the SAM-adenine moiety. The failure to detect any binding for compound 5155566 could be caused by the conformational plasticity of the binding pocket.

The third compound not displaying any binding was compound 4051493, which is based on a purine scaffold linked to a piperazine ring. This compound was predicted to form only four hydrogen bonds with the receptor (Fig. 4.8) and all of these are formed under participation of the purine-moiety. The piperazine moiety with its attached terminal groups does not participate in any interactions and the shape complementarity for this part of the ligand is poor. The combination of a small number of hydrogen bond interactions being formed and the poor shape complementarity can explain the lack of observable binding.

But apart from the challenges this project posed, one new ligand was discovered for this riboswitch. Compound 5669778 (Tab. 4.1) is the first ligand for the SAM-I riboswitch which is not a close SAM analog. The binding data presented here

(Fig. 4.10) still leaves some room for improvement and the measurements need to be repeated to obtain a statistical measurement of accuracy. While being a weak binder on an absolute scale the compound does exhibit a comparatively strong ligand efficiency (Hopkins et al., 2004). The 370 μM with 17 non-hydrogen atoms corresponds to a ligand efficiency of approximately 0.28 $\frac{kcal}{mol \cdot HA}$. As a comparison the ligand efficiency of the SAM is 0.33 $\frac{kcal}{mol \cdot HA}$ and 0.24 $\frac{kcal}{mol \cdot HA}$ for SAH. Therefore, the discovered ligand is not as good as SAM in terms of ligand efficiency but performs better than the SAM analogues known to bind the riboswitch. The classic target value for drug discovery 0.3 $\frac{kcal}{mol \cdot HA}$ is just slightly higher than the value obtained and in a drug discovery scenario this would still be considered an interesting binder. Also structural investigation of the binding mode would be desirable. It is crucial to confirm the observed binding mode of the compound is as predicted by DOCK 3.5.54 (Fig. 4.8 D). Nevertheless, this provides a first success in ligand discovery for the SAM-I riboswitch.

Additionally the presented data only represent a first attempt in structure based ligand discovery for the SAM-I riboswitch and the target is far from being thoroughly explored. Potential options include a virtual screening focused on the area of the binding site usually occupied by the SAM methionine. Also docking into the binding site configuration of the apo-structure is conceivable. A third option would be targeting the area of the binding pocket currently not occupied by any ligand atoms next to the methionine moiety of the SAM ligand (Fig. 4.11). Finally, it might be possible to utilize the identified compo9und as a building block and synthesize compounds featuring this basic core with potentially higher affinity.

The SAM-I riboswitch has been found in pathogens especially firmicutes including the *Bacillus* genus for which effective novel treatment options are urgently required. Compounds could be tested in the non-pathogenic model organism *B. subtilis* which contains eleven SAM-I riboswitches in its genome before studies with pathogenic organisms are carried out. The pathogenic species for potential

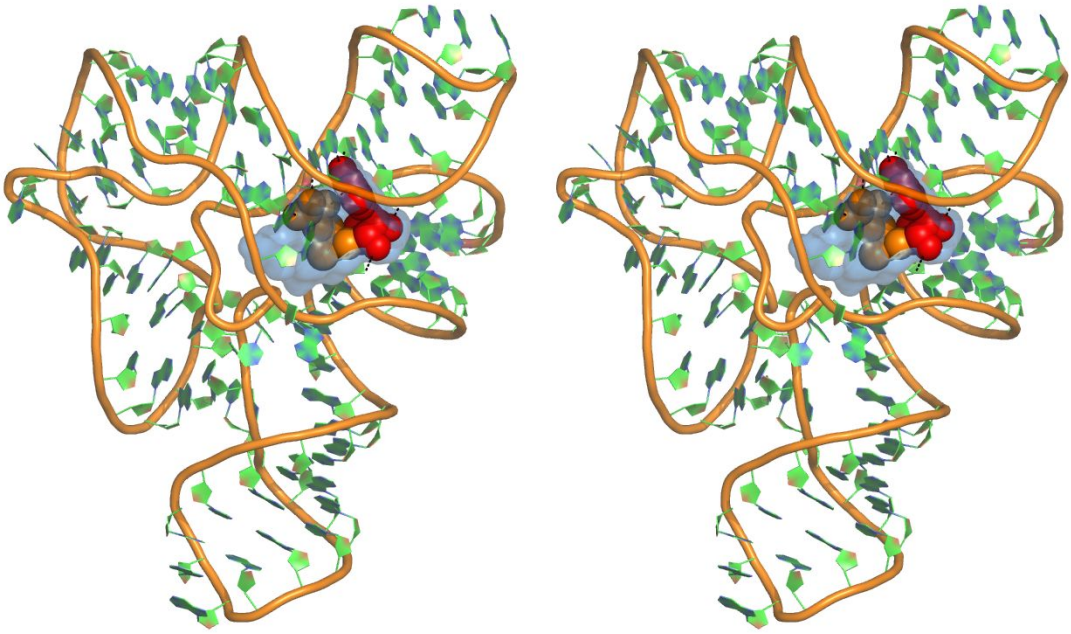


Figure 4.11: Wall-eyed stereo view of a schematic representation of the SAM-I riboswitch and its SAM binding pocket. The riboswitch is shown in cartoon representation. The SAM ligand atoms are shown as spheres, red spheres for the adenosine part of the ligand and amber spheres for the methionine part. The cavity making up the binding pocket is shown as the transparent blue surface.

clinical application include *B. anthracis*, *B. cereus* and *Listeria monocytogenes*. *L. monocytogenes* contains seven putative SAM-I riboswitches (Loh et al., 2009), and is one of the most common food borne pathogens. The SAM-I riboswitch would offer an opportunity of manipulating all genes regulated by these seven SAM-I riboswitches simultaneously.

On a broader scope successful virtual screening studies for RNA targets are still rare and the establishment of validated methods is vital towards development of drug discovery methodology for RNA targets in general.

Chapter 5

Stabilization of the k-turn in the SAM-I riboswitch by tertiary interactions

5.1 Results

5.1.1 Analysis of the SAM-I k-turn

The standard k-turn motif as identified from ribosomal structures (Klein et al., 2001) can be described by the minimal consensus shown in Fig. 5.1 A. The *T. tengcongensis* SAM-I riboswitch (see Fig. 5.1 B) contains a k-turn that matches this consensus perfectly. The k-turn is found in the P2 helix (Fig. 5.1 C; Montange and Batey 2006; Montange et al. 2010). The k-turn introduces an angle of about 60 ° in the P2 helix and places L2 in close proximity to J1/4 and J3/4, allowing the formation of the central pseudoknot. The k-turn contains the non-bulged strand residues comprising nucleotides 17-22 and the bulged strand nucleotides comprising nucleotides 30-38. Even though the k-turn is essential for folding of the riboswitch (Hennelly and Sanbonmatsu, 2010) and ligand binding, no direct interactions are formed between the SAM-ligand and the k-turn region

as can be seen in the structure of this riboswitch (Montange and Batey, 2006; Montange et al., 2010). The SAM ligand in the crystal structure is located more than 13 Å from the k-turn region (residues 17-22 and 30-38). In particular the two G-A base pairs at the 1n-1b/2n-2b positions are more than 15 Å distant from the SAM ligand. Consequently any influence of SAM binding on the k-turn structure must be mediated through tertiary interactions of the RNA structure. It appears that the k-turn provides a key element for the tertiary interactions

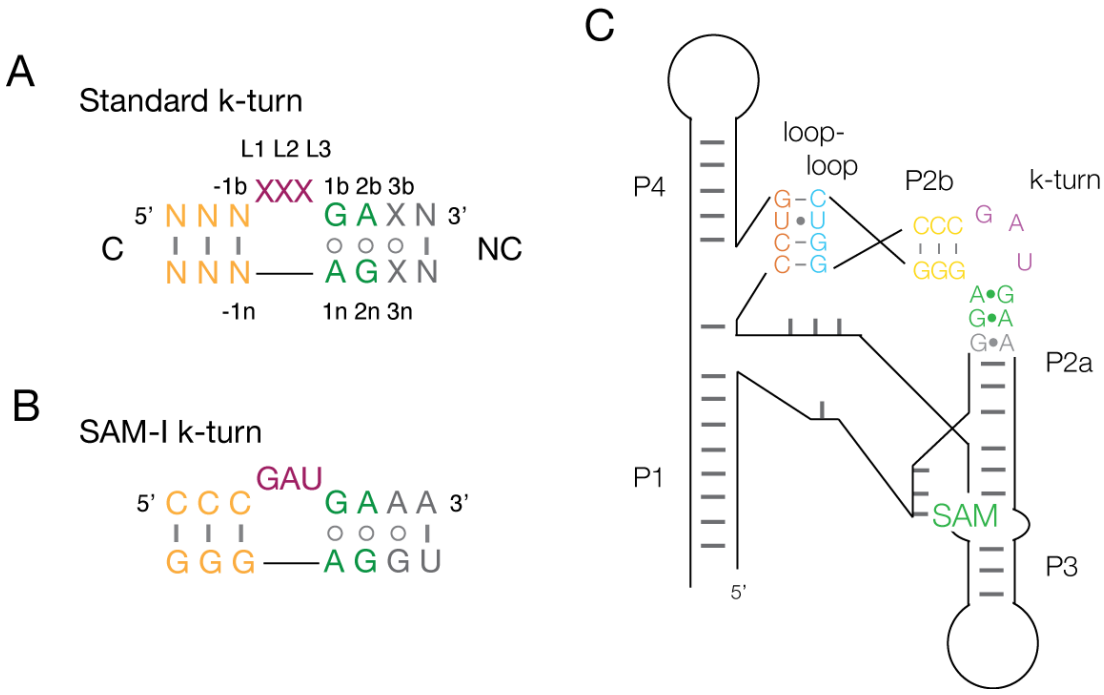


Figure 5.1: A: The consensus sequence of a standard k-turn. The individual nucleotides are labeled according to the naming convention described in Liu and Lilley 2007. B: The sequence of the SAM-I riboswitch k-turn C: Schematic drawing of the overall structure of the *T. tengcongensis* SAM-I riboswitch. Shown are the k-turn enabling the loop-loop interaction of P2 and the SAM-binding site. The grey bars indicate areas of double helix but do not correspond to individual base pairs.

that lead to the formation of the ligand binging site by positioning the L2 region. Also the k-turn is formed in the apostructure in the absence of ligand (Stoddard et al., 2010).

The near-consensus Kt-7 of *Haloarcula marismortui* is very similar in sequence to the SAM-I riboswitch k-turn. The SAM-I riboswitch comprises three Watson-

Crick base pairs at position of the -3 to -1 on the bulged and non-bulged strand. The 1 to 3 nucleotides of both strands form consecutive GA base pairs, the first two of which are extremely conserved in k-turns. The loop nucleotides consist of the sequence G-A-U in the wildtype riboswitch (PDB-file 2gis; Montange and Batey 2006) and G-A-C in the U34C/A94G mutant (PDB-file 3gx5). The U34C/A94G mutant gave a higher resolution crystal structure (2.4 Å; Montange et al. 2010) and therefore this sequence was adopted for subsequent analysis.

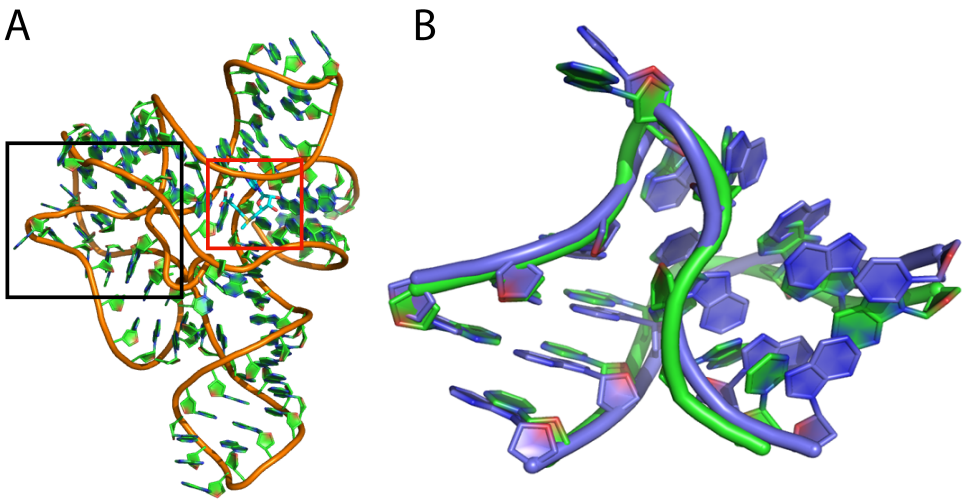


Figure 5.2: A: Overall structure of the SAM-I riboswitch (PDB-code 3gx5) in cartoon representation, the k-turn region in the black box, the SAM ligand (cyan carbon atoms) in the red box. B: Superposition of the *H. marismortui* kt-7 (green cartoon representation) and the SAM-I riboswitch k-turn (blue cartoon representation). The RMSD of the superposition of the two structures is 1.4 Å.

Since the nucleobase of C34,which is located in the L3 position of the k-turn, does not form any contacts within the riboswitch, this substitution is expected not to alter the riboswitch binding properties. In solution the L3 base moiety is pointing into bulk solvent and accordingly does not contribute any intramolecular interactions. It is the optimized crystal contacts of the U34C substitution which lead to better diffracting crystals by inter-molecular interactions, and thus this variant forms a favorable system to study k-turn folding in the context of an RNA molecule making important tertiary interactions.

The three-dimensional structures of *H. marismortui* Kt-7 (PDB file 3cc2; Blaha et al. 2008) and the SAM-I riboswitch k-turn superimpose with an RMSD of 1.4 Å (Fig. 5.2 B). A comparison of the structures shows the presence of all the key hydrogen bonds of a standard k-turn (Liu and Lilley, 2007). The G1b-A1n and A2b-G2n base pairs which have been shown to be of critical importance to k-turn structure, are conserved. Both are *trans* sugar edge - Hoogsteen edge base pairs and this interaction pattern is identical in Kt-7 and the SAM-I riboswitch k-turn. Also the L1 O2' interaction with A1n N1 which prevents k-turn folding if disrupted is fully conserved between both k-turns. Since it is known that near-consensus k-turns can fold in isolation (Matsumura et al., 2003; Goody et al., 2004; Schroeder and Lilley, 2009), it was expected that the SAM-I riboswitch k-turn would fold as an isolated duplex. The aim of the following experiments was to probe the sequence requirements of a kink-turn in isolation and in the context of a larger, structured RNA molecule to elucidate the possible role of tertiary interactions in k-turn folding.

5.1.2 The SAM-I K-turn in isolation

The experiments presented in this section were carried out by Kersten T. Schroeder. Since they form an integral and essential part of the whole project they are presented in the same detail as the other parts of the project. The SAM-I riboswitch k-turn is known to fold in the context of the full aptamer domain from structural data (Montange and Batey, 2006; Montange et al., 2010). However the behavior of the k-turn in isolation was unexplored at the outset of the study. To investigate this aspect, the SAM-I riboswitch k-turn sequence was incorporated in a double helical RNA molecule which was 5'-terminally labeled with fluorescein as a donor at one end and Cy-3 as an acceptor at the other (Fig. 5.3).

This construct allows the measurement of end to end distance and hence kinking of the RNA-axis via FRET. The FRET efficiency was measured using the acceptor

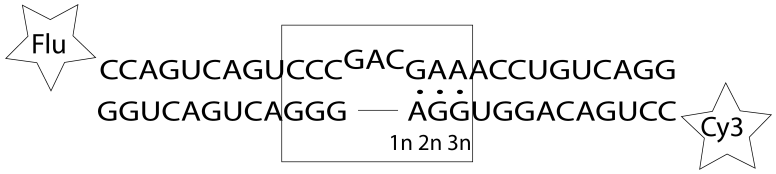


Figure 5.3: SAM-I riboswitch k-turn incorporated in a short RNA double helix used for FRET experiments to measure k-turn folding. Each strand is 5'-labeled with a fluorophore, either fluorescein or Cy3. Substitutions were made on the 1n-3n positions as described.

normalization method (Clegg et al., 1992). The value of E_{FRET} in the absence of metal ions is in good agreement with the expected value for a 3-base bulge (Gohlke et al., 1994). As previously demonstrated for Kt-7, the end to end distance decreases upon k-turn folding leading to an increased FRET efficiency. Addition of magnesium ions to the SAM-I k-turn in isolation results in a change of FRET efficiency $\Delta E_{FRET} = 0.32$, and the isotherm can be fitted to a simple two-state binding model with a magnesium half concentration $[Mg^{2+}]_{\frac{1}{2}} = 50\,\mu$M and a Hill coefficient of $n = 0.8$ (Fig. 5.4).

The triple substitution A1nC/G2nU/G3nU was investigated in isolation, which is expected to behave like a standard 3-base bulge without any k-turn folding occurring upon magnesium addition as the three GA pairs are each converted to Watson-Crick base pairs. Upon addition of magnesium ions no increase in FRET could be observed (Fig. 5.5).

Next the extent of variation compatible with k-turn folding was explored. The G2nU/G3nU variant underwent no folding in isolation (Fig. 5.5) similar to the triple substitution. In contrast the G3nU variant (Fig. 5.4) exhibited partially impaired folding, with $\Delta E_{FRET} = 0.19$, $[Mg^{2+}]_{\frac{1}{2}} = 110\,\mu$M and a Hill coefficient of $n = 0.9$. As all modifications to the 1n and 2n bases were deleterious to k-turn folding to this point, these positions were considered in more detail.

The An1C substitution gave no indication of folding (Fig. 5.4). Similarly the A1nU substitution (Fig. 5.5) and the A1nG substitution exhibited no indication of folding confirming that adenine is the only base in the 1n position that is

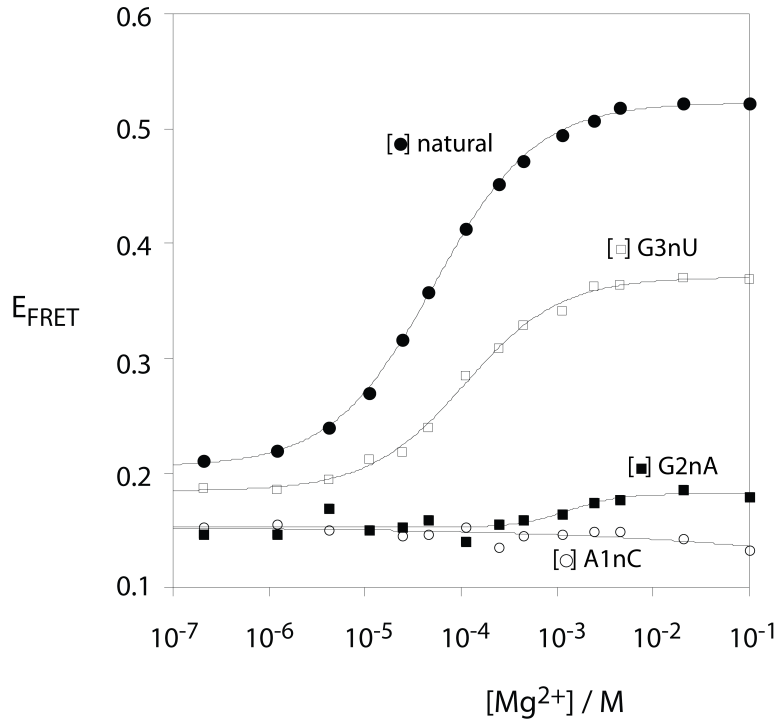


Figure 5.4: E_{FRET} of the k-turn constructs (Fig. 5.3) as a function of magnesium concentration as measured by steady state fluorescence spectroscopy. Data were fitted to a two state binding model. The natural sequence k-turn is represented by filled circles, the G2nU substitution by open squares, the A1nC substitution by open circles and the G2nA substituted k-turn by filled squares.

consistent with k-turn folding in isolation. Turning to the 2n position, all three single substitutions of the G2n resulted in no observable folding.

In summary, all substitutions tested in the two GA base pairs adjacent to the loop prevent k-turn folding in isolation while the G3nU mutation results in impaired, but measurable folding ability.

5.1.3 Folding of the SAM-I riboswitch k-turn in context of the riboswitch

The experiments presented here were conducted in collaboration with Kersten T. Schroeder. Since the k-turn seemed very sensitive to modification in isolation, it was investigated how this compares to the situation in context of the SAM-I riboswitch. From the data available it was hypothesized that k-turn folding is a

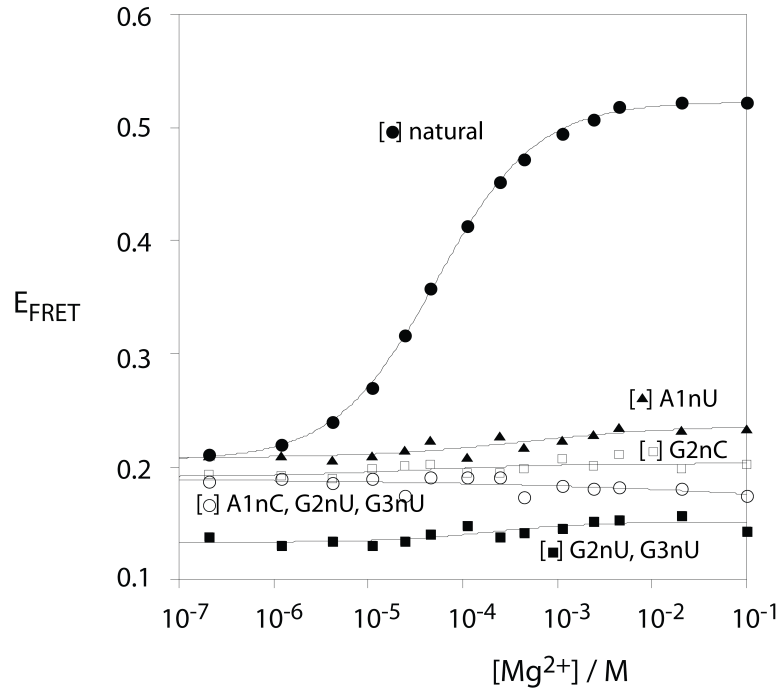


Figure 5.5: E_{FRET} of the k-turn constructs (Fig. 5.3) as a function of magnesium concentration as measured by steady state fluorescence spectroscopy. Data were fitted to a two state binding model. The natural sequence k-turn is represented by filled circles, the G2nC substitution by open squares, the A1nC/G2nU/G3nU triple substitution by open circles, the A1nU substitution by filled triangles and the G2nU/G3nU double substituted k-turn by filled squares.

key element in the global architecture of the riboswitch and therefore essential for ligand binding. ITC measurements were performed with the standard k-turn and the triple substitution G1nC,G2nU,G3nU to see if SAM binding might report on k-turn foldability. For the standard k-turn sequence a binding constant of $K_D = 0.5 \pm 0.3\,\mu$M was measured (Tab. 5.1, Fig. 5.6 A), in agreement with published data (Montange et al., 2010).

In contrast the triple mutant showed no heat evolution indicative of SAM binding (Fig. 5.6 B). The ablation of the k-turn has evidently prevented the riboswitch from adopting a conformation in which it can bind its SAM ligand. This suggests SAM-binding can be used as a measure of k-turn foldability as hypothesized.

Next the riboswitch with a G3nU substituted version of the k-turn was probed for SAM-binding and $K_D = 0.34 \pm 0.04\,\mu$M was measured (Tab. 5.1 C). Heat was

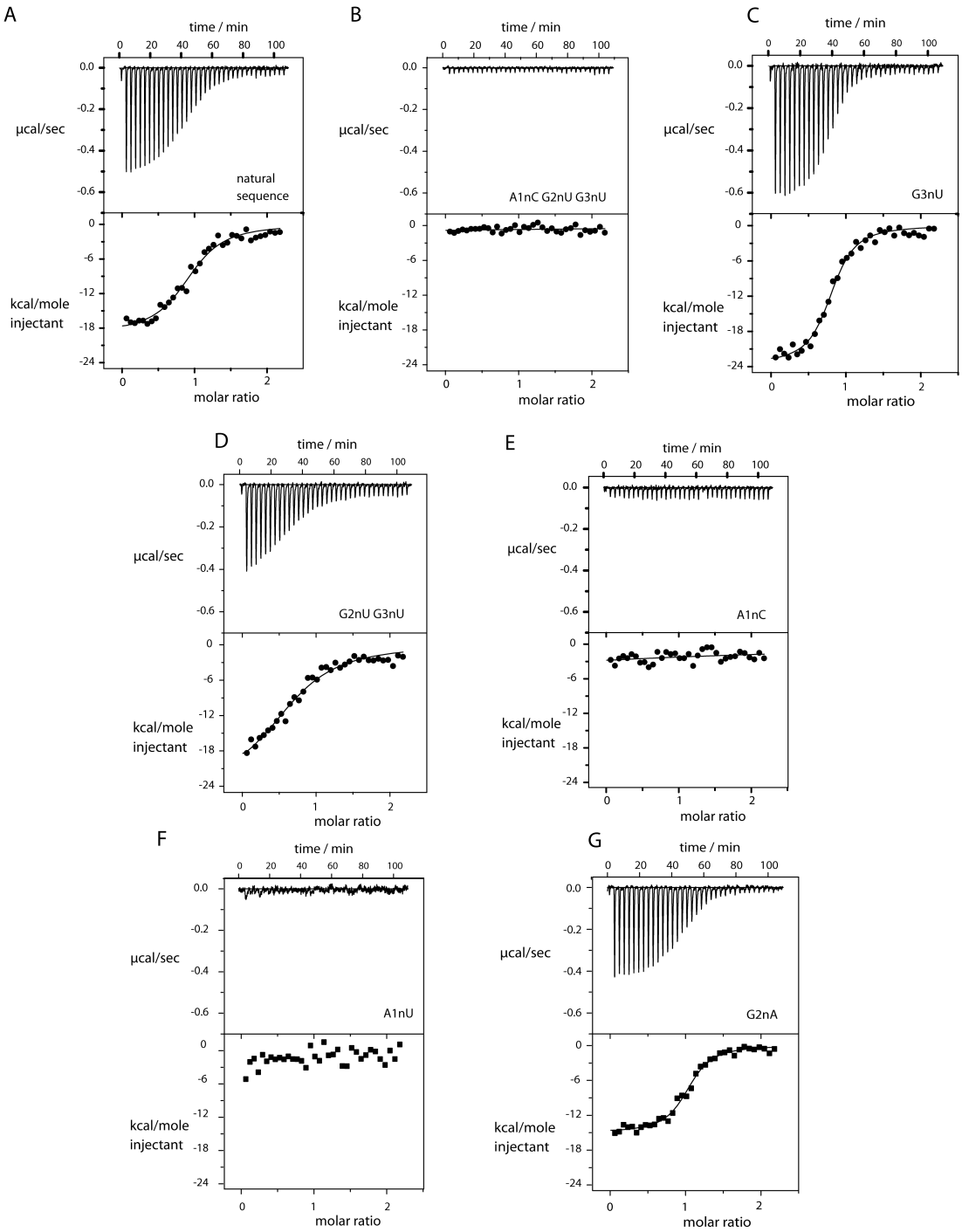


Figure 5.6: Isothermal titration calorimetry of SAM binding to the SAM-I riboswitch with the natural k-turn and modified k-turns. SAM is titrated into a solution of SAM-I riboswitch and the evolving heat is measured. Data are then fitted to a two-state binding model and thermodynamic parameters are summarized in Tab. 5.1.

Table 5.1: Thermodynamic binding parameters determined by ITC for SAM ligand titrated into a solution containing the SAM-I riboswitch. Data were fitted to a two state binding model resulting in the tabulated parameters.

k-turn	n	ΔG° ($\frac{kJ}{mol}$)	K_D (μM)
natural	0.85 ± 0.10	-36 ± 1	0.5 ± 0.3
A1nC/G2nU/G3nU	ND	ND	ND
G2nU/Gn3U	0.7 ± 0.1	-33.1 ± 0.4	2.0 ± 0.3
G3nU	0.74 ± 0.08	-37.6 ± 0.3	0.34 ± 0.04
A1nC	ND	ND	ND
A1nU	ND	ND	ND
G2nC	ND	ND	ND
G2nA	1.04 ± 0.08	-37.8 ± 0.5	0.31 ± 0.06

evolved and the measured affinity was identical within error to that obtained for the natural k-turn. A binding constant of $K_D = 2.0 \pm 0.3\,\mu$M was determined for the G2nU/G3nU variant (Fig. 5.6 D, Tab. 5.1), which is about fourfold weaker than the affinity of SAM to the riboswitch with the natural k-turn. A1nC (Fig. 5.6 E) and A1nU (Fig. 5.6 F) substitutions prevented measurable SAM binding to the riboswitch. The same result was obtained for the G2nC substitution but, surprisingly, the G2nA substitution exhibited SAM binding (Fig. 5.6 G) with a binding constant of $K_D = 0.31 \pm 0.06\,\mu$M which is comparable to the riboswitch with the standard k-turn sequence (Tab. 5.1).

Despite preventing folding of the isolated k-turn, the A2n-A2b basepair does not prevent SAM binding to the riboswitch, suggesting that the k-turn is folded normally. This indicates a stabilizing effect of the tertiary interactions within the aptamer domain on the k-turn structure. However, SAM-binding is a very indirect means of monitoring k-turn folding and the possibility that SAM binds by an alternative conformation of the k-turn cannot be excluded. Therefore, structural data were obtained to see if the k-turn region in these modified riboswitches is folded into the conventional three-dimensional motif.

5.1.4 Crystal structures of the SAM-I riboswitch with G2nA and G2nU/G3nU substitutions in the k-turn

The riboswitch variant bearing the G2nA mutation in the k-turn region was cocrystallised with SAM. Crystals were obtained in hanging drop format by mixing 1 μL of mother liquor with 1μ L of 400 μM RNA plus 1 mM SAM in 40 mM Na-cacodylate (pH 7.0). The mother liquor for the crystal used for data collection contained 10 mM $MgCl_2$, 40mM Na-cacodylate (pH 7.0), 12 mM spermine-HCl, 12 % (v/v) MPD and 80 mM KCl. The drops were seeded with a seed stock containing crystals of the unmodified RNA. The crystallisation conditions described above were very similar to those obtained in-house for the unmodified riboswitch and the crystals had the same appearance. These conditions were virtually identical with the published crystallisation conditions (Montange et al., 2010). Crystals were cryoprotected in mother liquor containing 30 %(v/v) MPD and mounted on cryo-loops followed by flash-freezing in liquid nitrogen.

Diffraction data were collected at the ESRF, with the best crystal diffracting to 2.6 Å. A complete data set was obtained, and analysed to solve the riboswitch structure bearing the G2nA mutation (Tab. 5.2).

The space group ($P4_32_12$) and unit cell dimensions ($a = b = 58.46$ Å, c=155.26 Å, $\alpha = \beta = \gamma = 90\,^\circ$) match those for the published structure of the unmodified riboswitch (PDB-code 3gx5) closely. This allowed an initial model for the modified riboswitch to be obtained by molecular replacement. This model was subsequently refined with the R-factors for the final model being $R_{work} = 20.0\,\%$ and $R_{free} = 25.9\,\%$.

The overall fold of the G2nA structure is identical to that of the unmodified riboswitch. An alignment based on the phosphorus atoms of all nucleotides of the unmodified riboswitch and the G2nA variant yields an RMSD of 0.74 Å. Electron density is well defined for the whole riboswitch and the SAM ligand.

The k-turn region (nucleotides 17-22 and 30-38) of natural and modified ribo-

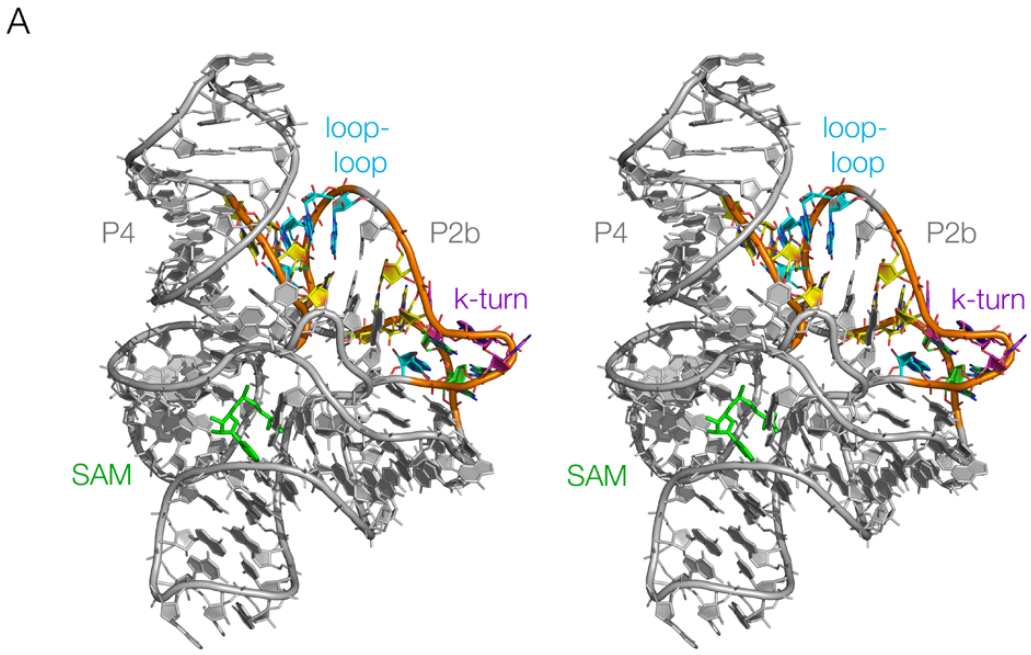


Figure 5.7: Wall-eyed stereo image of the SAM riboswitch structure with a G2nA substitution in the k-turn sequence. The structure is shown in grey cartoon representation except for the SAM-ligand shown in green sticks and the k-turn region and the loop-loop interaction which is shown in color.

switch align with an RMSD of 0.53 Å(Fig. 5.8). The k-turns do not differ significantly in their backbones. The $2F_o - F_c$ map shows that base and backbone regions are well defined by electron density for all residues (not shown). To exclude the possibility of model bias, a composite omit map was calculated for the final model using PHENIX (Adams et al., 2010). The backbone and most nucleobases are also defined by the obtained composite omit electron density map of the k-turn region (Fig. 5.9). It can therefore be concluded that the presented model is not a result of model bias. Additionally the structure was validated with MOLPROBITY (Chen et al., 2010) and performed comparably to the unmodified riboswitch structure (Montange and Batey, 2006) The G2nA mutation creates an A-A basepair at the 2n-2b position which differs in its hydrogen bonding pattern from the G-A pair in the unsubstituted riboswitch. A single hydrogen bond from A2nN3 to A2bN6 is formed. A2n also donates a hydrogen bond from N1 to O3'

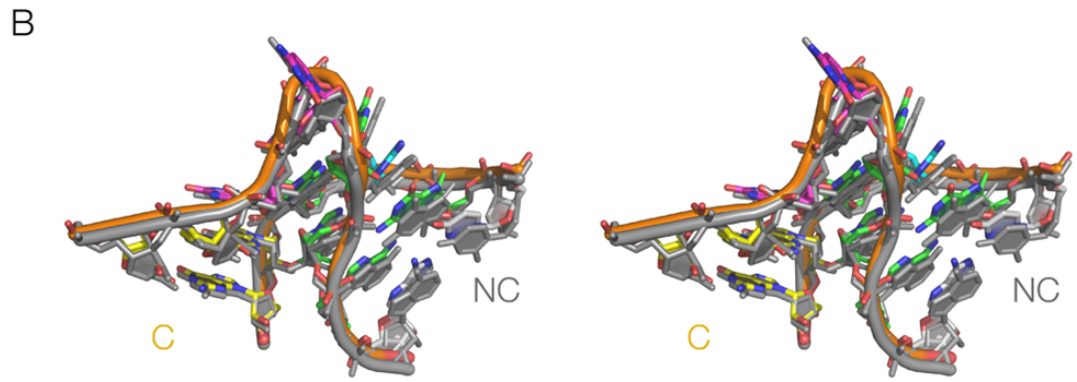


Figure 5.8: Wall-eyed stereo image of the k-turn from the unsubstituted SAM-I riboswitch (grey) and the G2nA substituted variant (color). Both structures exhibit the conventional k-turn fold, confirming that the G2nA substituted k-turn fold is indeed salvaged by tertiary interactions.

of G1b. These two hydrogen bonds appear to lock the A2n into the observed position, aided by base stacking of A2n between G18 and G35. This k-turn is the first simple k-turn with such an A-A pair for which a structure has been determined. Interestingly the *Thelohania solenopsae* Kt-23 has an A2n-A2b basepair in its natural sequence, although no structure is currently available.

In addition to the G2nA mutant, the crystal structure of the SAM-I riboswitch with a k-turn sequence having a G2nU/G3nU double substitution was determined. This k-turn sequence was shown not to fold in isolation like the G2nA substitution, but SAM binding to the riboswitch was observable, although impaired in context of the riboswitch (see above). Crystals were obtained using the same method as for the G2nA substituted variant, and the drop containing the crystal used for data collection contained 5 mM $BaCl_2$, 40 mM Na-Cacodylate (pH 7.0), 80bmM KCl, 12bmM spermine-HCl, 12 % MPD. Crystals were cryoprotected with mother liquor plus 15 % PEG. A crystal was mounted on a cryoloop and diffracted to 2.9 Å. Completeness was low in the highest resolution data shell (64 % between 2.9 and 3 Å), so the crystal structure is regarded as of 3 Å resolution.

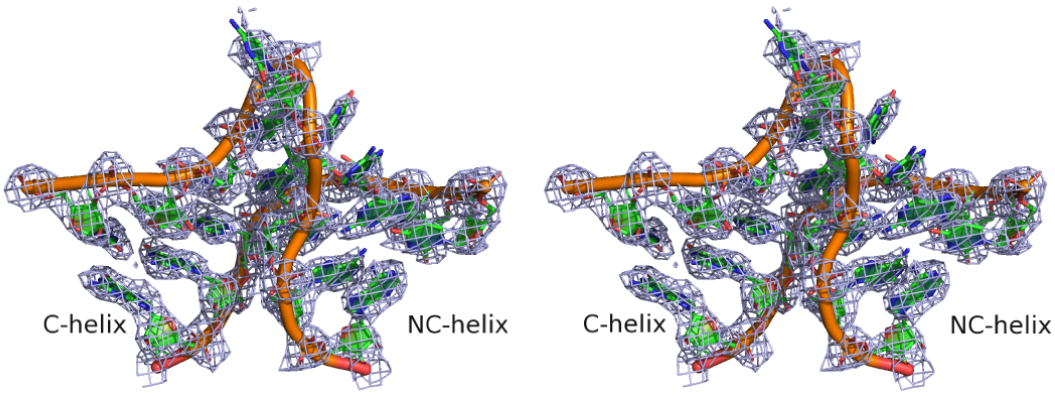


Figure 5.9: Wall-eyed stereo image of the k-turn region of the SAM-I riboswitch with the G2nA substitution. The RNA is shown in cartoon representation and the composite omit electron density map at 1 σ is shown in grey. The k-turn model is well defined by the electron density.

The space group ($P4_32_12$) was the same as for the unmodified riboswitch and the unit cell dimensions are virtually identical. After molecular replacement and refinement the final model (R_{work} = 26.9 %/R_{free} = 32.2 %) showed that the overall fold of the riboswitch does not change compared to the unmodified sequence. An RMSD of <0.5 Å is obtained upon alignment of the substituted and original structure based on the phosphorus atoms of all nucleotides. In general the riboswitch is well defined by electron density and the ligand is well defined for the adenosine moiety but lacks observable density for the methionine moiety. A composite omit map calculated for the final model confirmed the model free of potential bias associated with the $2F_o - F_c$-maps. The omit map at 1 σ-level around the k-turn region is shown in figure 5.10; it defines the sugar-phosphate backbone reasonably well but lacks density for many base moieties. Nevertheless it can be concluded that the model represents the diffraction data well. Also when the structure was validated using MOLPROBITY (Chen et al., 2010) it performed comparable to the structure of the unmodified riboswitch Montange

and Batey (2006)

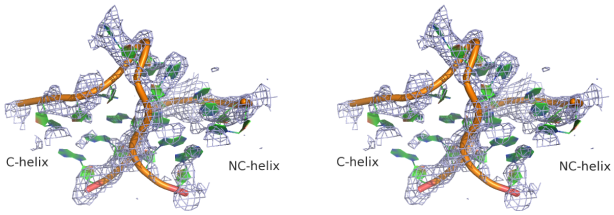


Figure 5.10: Wall-eyed stereo image of the k-turn region of the SAM-I riboswitch with the G2nU, G3nU substitution. The RNA is shown in cartoon representation and the composite omit electron density map at 1 σ is shown in grey. While the nucleobases are not well defined by the electron density, the backbone, in particular the phosphate groups, defines the overall conformation of the k-turn beyond doubt.

Looking at the k-turn region, an alignment (residues 17-22 and 30-38) based on the phosphorus atoms, which are regarded to be the best defined backbone atoms at this resolution yields an RMSD of 0.75 Å. In conclusion the k-turn adopts the conventional fold in context of the determined crystal structure despite the impaired interactions of the 2n-2b/3n-3b base pairs.

5.2 Discussion

5.2.1 Folding of the k-turn in isolation

In this study, it was investigated how k-turns react to sequence alterations with respect to their molecular environment. The folding of the k-turn motif in solution, one needs to be aware, involves a two-state dynamic equilibrium between folded and unfolded states. The balance of this equilibrium is determined by the relative energies associated with these states, which depends on parameters including ionic strength, temperature and pH. In low ionic strength the repulsion between juxtaposed backbone atoms of the k-turn raises the energy of the folded state above that of the unfolded state. Consequently, very few molecules are found in the folded state which is therefore not observable by steady-state

FRET. As the ionic strength rises, the repulsion between the backbone atoms is shielded by diffuse counter ions and the energy of the folded state is lowered. This raises the population of the folded state. In the FRET data this is reflected in the smooth transition of low to high FRET efficiency. The natural sequence SAM-I k-turn is a good example for this situation.

Alterations of the sequence of the k-turn changes the relative energies of the folded and unfolded states. In some cases this results in a situation where the folded states are never favored compared to the unfavored states resulting in a loss of observable folding. An obvious example for this effect is the triple substitution A1nC/G2nU/G3nU variant. The three Watson-Crick base pairs that result, lead to a tight duplex with a three base bulge. The energy of the folded k-turn is elevated to such a level that it will have virtually zero population.

In principle, any substitution can alter the energetic balance either by altering the energy of the unfolded state or the folded state. As described for the triple substitution A1nC/G2nU/G3nU a combination of both is possible. The further substitutions described in the results section now probe to which extent the alteration of the k-turn motif is compatible with folding by increase of ionic strength. As described in detail in the results section almost all changes to the k-turn sequence were deleterious for k-turn folding in isolation. This is due to the selection of substitutions which probed the critical nucleotides for k-turn formation. The 1n and the 2n nucleotides do not tolerate any substitution without complete obliteration of k-turn folding. This agrees with the data published for the *H. marismortui* Kt-7 (Goody et al., 2004), and in context of the critical interactions identified by substitution with modified nucleotides (Liu and Lilley, 2007; Turner and Lilley, 2008).

The G3nU substituted k-turn exhibits a particularly interesting behavior. Folding is still observed with increase of ionic strength but the change in E_{FRET} is smaller and the FRET efficiency at high ionic strength is reduced compared to the unmodified k-turn. This can be explained by the model for k-turn folding

outlined above. At low ionic strength the unfolded state is energetically favored and thus low E_{FRET} is observed overall. As the ionic strength increases the energetic balance shifts, and the now more favorable folded states dominate. However the folded states are of slightly higher energy than for the unsubstituted k-turn, so that even when the shielding effect is at full capacity a larger residual fraction of unfolded states is present. It is important to observe that the $[Mg]_{\frac{1}{2}}$ of the G3nU substituted k-turn is not different from the unsubstituted k-turn, which indicates the substitution does not affect the shielding effect of the ions. The different end points of the titration curves indicate an incomplete transition from mostly unfolded to mostly folded states.

An alternative interpretation is the assumption that in the substituted k-turn the folded states are geometrically different from the unsubstituted k-turn, for example by an altered angle between the helical ends of the RNA. This would also result in an increased average distance between the FRET partners and therefore to a decrease in FRET efficiency. The identity of the transition point in the substituted and unsubstituted k-turn however, supports the model of incomplete transition.

5.2.2 k-turn folding in context of the SAM-I riboswitch

The situation which has to be considered for the interpretation of the ITC data is quite different from the situation described above. First, under ITC assay conditions the unmodified k-turn in isolation would be fully folded. Second, the readout of k-turn folding is SAM binding which is primarily dependent on the ability of the motif to fold in context of the riboswitch. Data indicating the importance of the k-turn for ligand binding has been published (Hennelly and Sanbonmatsu, 2010). As outlined earlier the SAM-I riboswitch has been crystallized in the presence and absence of ligand (Montange and Batey, 2006; Montange et al., 2010; Stoddard et al., 2010). Therefore it can be concluded that the folded conformation is

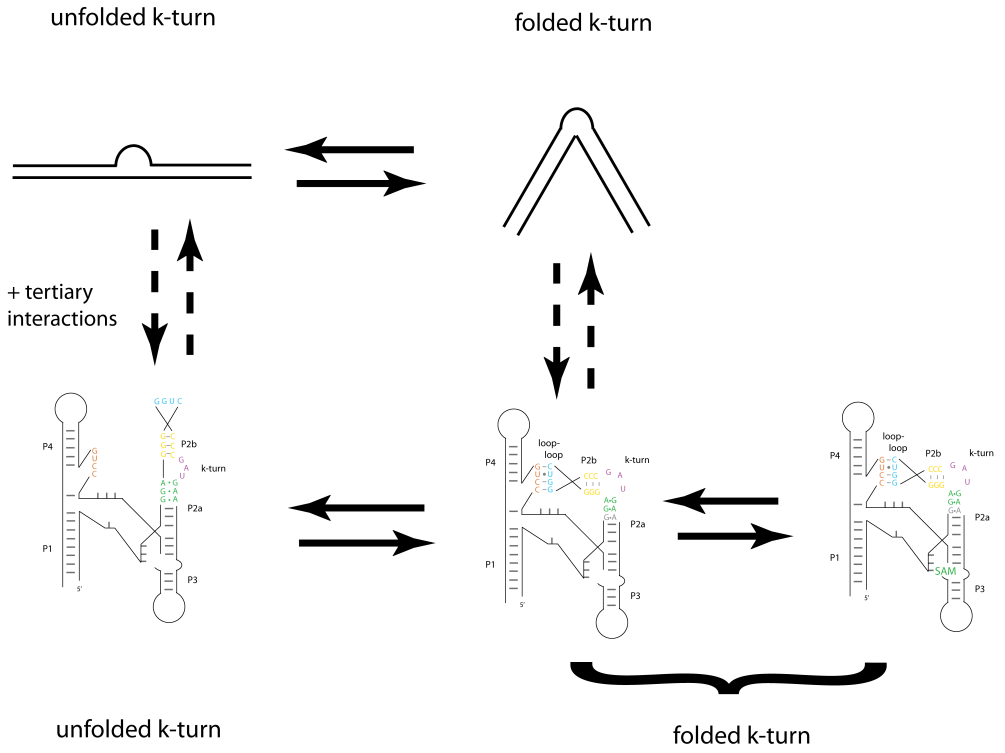


Figure 5.11: Schematic diagram of k-turn folding in isolation and in context of the SAM-I riboswitch. In isolation the folded and unfolded k-turn are in an dynamic two state equilibrium, with the relative occupancies for both states determined by external parameters such as ionic strength. In case of the riboswitch a three state equilibrium exists, the RNA in the ligand unbound form also is in a dynamic equilibrium of states with folded and unfolded k-turns, but usually this equilibrium is strongly dominated by the states with a folded k-turn. The states with the folded k-turn are then able to bind SAM and SAM bound and unbound states are in a second equilibrium. SAM binding locks the k-turn in its folded state and stabilizes the molecule.

stable in solution in the absence of the ligand, at least for the isolated aptamer domain. The situation in context of the aptamer domain and expression platform might very well be different. The expression platform is expected to introduce competition between alternative structures of the overall riboswitch but no data are available for this case. Under the conditions of the ITC experiment, a certain fraction of the aptamer domain will adopt conformations resembling that of the crystallographically-determined apostructure which has the k-turn folded in the absence of SAM ligand. This structure can bind SAM resulting in the evolution of heat and a measurable binding curve. When the SAM-I aptamer is modified

to disrupt k-turn folding, the SAM-binding conformation is inaccessible and no binding is observed. The difference in behavior for k-turns in isolation and in the context of the riboswitch stems from the additional contribution of tertiary interactions to overall conformational energy. While for the isolated k-turn a small change in the energy of the folded or unfolded states can tip the balance to obliterate folding, in context of the riboswitch with its additional contributions to folding the overall balance is expected to be more robust. In particular the energy associated with the loop loop interaction of the P2 to the four nucleotides in P4 may drive the k-turn towards the folded state.

The riboswitch bearing the unmodified k-turn of course folds and binds ligand, and serves now as the standard for the behavior of riboswitches with modified aptamers. The triple-substituted riboswitch does not exhibit any SAM binding, indicative of disruption of k-turn folding. The SAM-binding conformation is not accessible and thus disruption of the k-turn cannot be salvaged by the tertiary interactions within the RNA molecule. Interestingly the same is true for the A1nC and the A1nU-substituted variant. This supports the importance of the A1n-G1b base pair for the stability of the k-turn fold, in agreement with previously reported data (Goody et al., 2004; Turner and Lilley, 2008).

In case of the G2nA substitution the energetic balance is in favor of the folded riboswitch since SAM binding deduced from the ITC is unchanged. Even though the FRET experiments show that the folded state of the k-turn is not stable in isolation, the tertiary interactions within the RNA can compensate and the overall folding of the riboswitch is not measurably perturbed. The SAM ligand can then bind to the aptamer domain similarly to the unmodified RNA. The same analysis is true for G2nU/G3nU doubly substituted riboswitch. However, a reduced affinity of the SAM ligand is observed. This can be explained by a stabilization effect of the SAM ligand for the folded states. The energy of the tertiary interactions is not sufficient to stabilize fully the folded conformation in the absence of SAM. This could be demonstrated by footprinting experiments

to determine the fraction of fully folded molecules. Therefore, part of the SAM binding energy needs to compensate and stabilize the folded state, resulting in a decreased net binding energy which is directly linked to SAM affinity.

5.2.3 Crystal structures of the SAM-I riboswitch with modified k-turns

There is a possible alternative interpretation of SAM binding data for the G2nA and the G2nU/G3nU variants invoking a SAM-binding conformation that might exist without a folded k-turn. To test this possibility and to dissect further the modified interactions in the substituted k-turns, crystal structures for both riboswitch variants were obtained. The data for the G2nA-substituted variant show that the conventional k-turn fold is conserved, preserving the A-minor interactions and most of the standard hydrogen bonds. The computed composite omit map demonstrates the validity of the crystallographic model beyond potential model bias. The G2nA substitution results in an overall loss of one hydrogen bond but the nucleobase still maintains contacts to A2b as well as the 2' hydroxyl group of G1b. At the same time the A2n cannot form any favorable alternative interactions with A2b. In conclusion the substitution probably leads to a slight elevation of the energy of the folded state while the energy of the unfolded state is not markedly altered. The loss of energy associated with this hydrogen bond can then be compensated for by the overall folding energy of the riboswitch.
For the G2nU/G3nU substituted riboswitch the data are not of comparable quality. A resolution of 2.9 Å in combination with low completeness in the highest resolution shell led to electron density maps that reveal much less detail. Nevertheless, the overall fold of the riboswitch and the k-turn could be confirmed. The composite omit map allowed tracing of the backbone at least while the orientations of individual nucleobases remained elusive. In conclusion the data certainly confirm the conventional k-turn fold for the G2nU/G3nU substituted riboswitch,

but a more detailed analysis of the contributions of individual intramolecular interactions to the behavior of this variant is not justified by the available data.

5.2.4 Conclusions

The data presented suggest a delicate energetic balance of folded and unfolded states for the natural SAM-I k-turn in isolation. This balance is easily perturbed by substitutions in the key bases defining the k-turn motif with major effects on folding. However, in the context of a larger RNA-molecule such as the SAM-I riboswitch, folding is in some cases maintained by the contribution of intramolecular tertiary interactions. This demonstrates that, in addition to protein binding and high ion concentration, tertiary interactions provide a third mechanism for the stabilization of the conventional k-turn fold.

However, several open questions remain and further experiments are needed to understand fully the folding of the k-turn motif. Single molecule studies have contributed considerably to the understanding of RNA folding in several cases (Lemay et al., 2006; Ouellet et al., 2010; Neupane et al., 2011). Recently, a study has applied single molecule measurements to the SAM-I riboswitch (Heppell et al., 2011), but this study did not look at the k-turn folding in any detail. Therefore, applying single molecule studies to the SAM-I riboswitch k-turn both in isolation and in context of the riboswitch would elucidate the k-turn folding further.

Furthermore, it remains unknown whether in principle any real k-turn sequence is compatible with riboswitch folding. Or phrased differently, can the SAM-I k-turn be replaced by k-turn sequences from the ribosome of different species. This leads to the question whether different k-turns are slightly different in terms of their kink-angle or in terms of the introduced helix torsion. Most of the ribosomal crystal structures from which k-turn structures were derived have a resolution > 2.5 Å and are structures of very large complexes (Schluenzen et al., 2000; Klein et al., 2001; Weixlbaumer et al., 2008). Consequently individual k-turns

might not be optimally modeled, or small differences in structure might have been missed. It is conceivable that there are classes of distinct k-turn geometry. If this were the case k-turns of the same class would be expected to be interchangeable in a riboswitch context, while k-turns of a different class from that found in the original riboswitch might prevent folding.

Chemically synthetic RNA for the full length riboswitch is currently becoming available from commercial suppliers. This will enable dissection of the contribution of individual hydrogen bonds to k-turn folding in context of the full aptamer domain, similar to what has previously been done with *H. marismortui* Kt-7 in isolation (Liu and Lilley, 2007).

Finally, the identification of tertiary interactions as a folding factor for k-turns might be important for the assembly of the ribosome itself, since many ribosomal k-turns found in biological context deviate from the consensus sequence. In case of the *T. solenopsae* Kt-23 from the 16S rRNA, the A2n-A2b basepair occurs naturally. The correct k-turn fold can in this situation be driven either by the binding of the ribosomal proteins or by the tertiary interactions within the rRNA. This has implications on the possible mechanisms of ribosome biogenesis. Either the tertiary interactions drive the k-turn folding which then allows protein binding or vice versa. Since k-turns are found multiple times in the ribosome structure it can be assumed that they are an important structural feature of ribosomes. A better understanding of the k-turn motif and the parameters vital to its folding will to contribute to a better understanding of the ribosome and its biogenesis.

Table 5.2: Statistics on data collection and refinement for the SAM-I riboswitch structures with modified k-turn sequences.

Sequence	G2nA	G2nU/G3nU
PDB code	2ygh	2ydh
Space group	$P4_32_12$	$P4_32_12$
Unit cell dimensions (Å)	a=b=58.64 c=155.26 $\alpha = \beta = \gamma = 90°$	a=b=61.89 c=157.63 $\alpha = \beta = \gamma = 90°$
Resolution range (Å)[1]	51.8-2.6 (2.74-2.6)	50-2.9 (2.0-2.9)
Observations	82286	41985
Unique observations	8929	6875
Completeness (%)	100.0 (100.0)	93.1 (64.5)
$\left\langle \frac{I}{\sigma(I)} \right\rangle$	22.6 (6.0)	40.6 (2.4)
$R_{merge}(\%)^2$	7.0 (38.3)	5.1 (42.3)
Multiplicity	9.2 (9.5)	6.1 (5.6)
Refinement statistics		
Resolution range (Å)	51.8-2.6	50-2.9
$R_{work}/R_{free}(\%)^3$	20.0/25.9	26.9/32.2
Number of atoms RNA/ligand	2033/27	2024/27
Number of atoms ions/water	6/19	5/2
Mean B-factor RNA/ligand	28.8/21.3	96.9/77.3
Mean B-factor ions/water	25.0/17.0	161.7/30
RMS bond length deviation (Å)	0.01	0.01
RMS bond angle deviation (°)	1.892	1.914

[1]Values in brackets are for the highest resolution shell.
[2]$R_{merge} = \sum (|I - \langle I \rangle| / I)$
[3]R-factor=$\sum |F_o - F_c| / \sum F_o$

Chapter 6

Concluding remarks and future perspective

6.1 Conclusion

All results presented and discussed above focus around the central question of structure and ligand recognition in riboswitches. While the different chapters focus on different aspects of this central topic one central theme seems to be common to all three of them: Individual interactions on a molecular level are generally understood and can be explained from theory, however their interplay in a given molecular scenario and the resulting consequences are far less understood.

6.1.1 Examples from the GRA

In the case of the purine riboswitch the importance of a water molecule for ligand binding was revealed. While binding of water molecules to biological macromolecules and the concept of water-mediated receptor ligand interactions are not difficult in themselves, treatment of these effects in a docking scenario is still a challenge in the field (Schneider, 2010). The fine balance between energetic cost and benefit of water binding and the identification of potential locations for

such water molecules are the subject of intense investigations.

The correct treatment of protonation and deprotonation of functional groups from both ligand and receptor is a second example for lack of understanding of the interplay between different effects. While the concept of pKa values is one of the most basic principles in chemistry the prediction of these pKa values and their variability in different environments for a given molecule is still challenging (Klebe, 2006). Recently more sophisticated models have been proposed and applied (Liao and Nicklaus, 2009; Kalliokoski et al., 2009; Manchester et al., 2010).

6.1.2 Examples from the SAM-I riboswitch

A problem similar to pKa prediction was aqueous solubility of the compounds selected from the virtual screening of the SAM-I riboswitch. Even though the principle factors influencing solubility are known, prediction of solubility for a given compound in a given solvent is still not an easy task. Even though predictors are available, they are far from reliable and aqueous solubility is usually approximated by the partition coefficient between n-octanol and water when solubility is assessed for a large number of compounds.

While the effects on the results of the ligand screening for the SAM-I riboswitch cannot be quantified, binding site flexibility was shown to be an important feature of the SAM-I riboswitch (Stoddard et al., 2010). Induced fit binding mechanisms and rearrangement of residues in the binding site are a well known challenge in docking (Yuriev et al., 2011; B-Rao et al., 2009). Efforts to include receptor flexibility into docking algorithms have a long history but success has been limited and research in this field is ongoing. A comprehensive understanding of RNA folding to a degree that would allowed incorporation of RNA flexibility into the prediction methods for ligand binding is still elusive.

In summary, for all cases mentioned above the individual contributions are under-

stood and can be rationalized but their interplay in a complex molecular scenario cannot be predicted. Seen from a different angle experimental results are rationalized readily using fundamental principles but prediction of experimental results from these very principles is challenging. However, the success of molecular docking demonstrated above shows that at least for some scenarios a qualitative and to some degree even quantitative understanding of ligand binding can be achieved. The utilized scoring function with its terms for electrostatic, van der Waals and desolvation energy was successful in approximating the energetics of binding and predicting the experimentally observed behavior. Interestingly, all the problems encountered in RNA-ligand docking are well known from protein-ligand docking and no RNA specific issues were identified in the chosen systems.

The results for the SAM-I k-turn in isolation and in the context of the riboswitch show that obtained data can be interpreted at a high level of detail. The k-turn offers the opportunity to study folding of a structural motif at the level of individual molecular interactions in a variety of environments. Quantification of the contribution of individual hydrogen bonds in different molecular environments seems feasible. The experiments performed demonstrate the power of combining quantitative folding assays with X-ray crystallography to understand the structure determining factors of the k-tun motif.

6.2 Long term perspective

The main challenge for molecular modeling and the detailed investigation of molecular systems in general will be to accurately include the interplay of different molecular effects and balance their individual contributions in order to predict the behavior of the system under a given set of parameters. In the past attempts to do so have been only of limited success and one problem observed was that with every additional effect considered new problematic situations appear. This was described as an "endless staircase" (Schneider, 2010) for molecular docking.

One way forward seems to perform detailed analyses of a single effect in a well defined model system to dissect the different contributions in a known setting. From there gradual generalization can be attempted. It remains to be seen whether this approach is feasible for all cases and weather some of the challenges will just prove to be too hard to tackle. The other avenue of progress apparent is to use the rapidly evolving computational technology and use more computational resources for the problem at hand. Originally, docking algorithms used very simplified models of ligand binding. The consideration of flexibility was simply not within reach of the computational resources. Recently, techniques and algorithms much more computationally demanding, such as molecular dynamics simulations, have been integrated into docking protocols (Marco and Gago, 2007). Again time will have to show whether simply increasing the computational resources will lead to a better overall performance of the predictors.

6.3 Short term goals and next steps

Finally the question remains, where to continue with the research presented above. For the docking part of the results, it can be said that both riboswitches used have not been exhausted as model systems for RNA-ligand docking. The purine riboswitch has so far only been looked at for the U74 version and whether the C74 version will yield comparable results is yet to be determined. Preliminary data on the C74 variant suggests that the docking approach will perform similarly. The purine riboswitch with its relatively small binding site would also be suitable for a more detailed retrospective study testing and validating improvements in the treatment of pKa effects and tautomerization. Docking into the receptor conformation observed in the 2'-deoxyguanosine bound structure (Edwards and Batey, 2009) is an attractive option as well. The enlarged binding site would allow exploration of a larger chemical space in the context of a known system.

For the SAM-I riboswitch further docking could be performed varying several parameters, including docking into the apostructure, and focusing on the area of the binding pocket usually binding the methionine moiety of the SAM ligand. The optimization of the docking protocol to include the full conformational space for the ligand could use the SFG and SAH bound structures of this riboswitch as a starting point.

For the aspect of k-turn folding three main avenues of interest can be identified:

1. Investigation whether the k-turn in the SAM-I riboswitch can be replaced by other k-turns, such as the ones found in the ribosome.

2. After identification of three main contributors of k-turn folding (intrinsic foldability, protein binding and tertiary interactions) it would be interesting to see how the different factors are of different importance for different k-turns in their native environment.

3. Use of synthetic full length SAM-I riboswitch including modifications at key residues to dissect the key interactions of the k-turn in the context of the full riboswitch.

In summary the systems chosen for experimental and computational investigation have helped to further the understanding of molecular recognition and folding in riboswitches as well as RNA in general. For most of the problems encountered starting points for optimization and many experiments to be performed in the future could be identified.

Appendix A

List of Abbreviations

- A adenine
- AR *V. vulnificus* adenine riboswitch
- AUC area under curve
- C cytidine
- DMSO Dimethyl sulfoxide
- DNA deoxyribonucleic acid
- DTT dithiothreitol
- EDTA ethylenediamine tetraacetic acid
- ESRF European synchrotron radiation facility
- FMN flavin mononucleotide
- FRET fluorescence resonance energy transfer
- GMP guanosinemonophosphate
- GRA *xpt-pbuX* guanine riboswitch C74U mutant from *Bacillus subtilis*
- G guanine
- HCV hepatitis C virus
- HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
- HIV human immunodeficiency virus

- HPLC high performance liquid chromatography
- ITC isothermal titration calorimetry
- MPD 2-methyl-2,4-pentanediol
- mRNA messenger RNA
- NTP nucleoside triphosphate
- PCR Polymerase chain reaction
- PDB protein databank
- PEG polyethylene glycol
- RMS root mean square
- RMSD root mean square deviation
- RNA ribonucleic acid
- ROC receiver operating characteristic
- rRNA ribosomal RNA
- SAC S-adenosylcystein
- SAH S-adenosylhomocystein
- SAM S-adenosylmethionine
- SFG sinefungin
- snoRNA small nucleolar RNA
- snRNA small nuclear RNA
- T thymine
- TAR transactivation response element
- t-BDMS *tert*-butyldimethylsilyl
- TPP thiamine pyrophosphate
- Tris 2-amino-hydroxymethyl-propane-1,3-diol

- tRNA transfer RNA
- U uracil
- UV ultra violet

Bibliography

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2):213–21.

Agrafiotis, D. K., Gibbs, A. C., Zhu, F., Izrailev, S., and Martin, E. (2007). Conformational sampling of bioactive molecules: a comparative study. *Journal of chemical information and modeling*, 47(3):1067–86.

Ashfaq, U. a., Javed, T., Rehman, S., Nawaz, Z., and Riazuddin, S. (2011). An overview of HCV molecular biology, replication and immune responses. *Virology journal*, 8(1):161.

Autexier, C. and Lue, N. F. (2006). The structure and function of telomerase reverse transcriptase. *Annual review of biochemistry*, 75:493–517.

B-Rao, C., Subramanian, J., and Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug discovery today*, 14(7-8):394–400.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science (New York, N.Y.)*, 289(5481):905–20.

Barbault, F., Zhang, L., Zhang, L., and Fan, B. T. (2006). Parametrization of a specific free energy function for automated docking against RNA targets using neural networks. *Chemometrics and Intelligent Laboratory Systems*, 82(1-2):269–275.

Barrick, J. E. and Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome biology*, 8(11):R239.

Batey, R. T., Gilbert, S. D., and Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, 432(7015):411–5.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4):271–81.

Bayne, E. H. and Allshire, R. C. (2005). RNA-directed transcriptional gene silencing in mammals. *Trends in genetics : TIG*, 21(7):370–3.

Beaucage, S. L. and Caruthers, M. H. (1981). Deoxynucleoside phosphoramiditesA new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, 22(20):1859–1862.

Bissantz, C., Kuhn, B., and Stahl, M. (2010). A medicinal chemist's guide to molecular interactions. *Journal of medicinal chemistry*, 53(14):5061–84.

Blaha, G., Gürel, G., Schroeder, S. J., Moore, P. B., and Steitz, T. a. (2008). Mutations outside the anisomycin-binding site can make ribosomes drug-resistant. *Journal of molecular biology*, 379(3):505–19.

Blouin, S. and Lafontaine, D. a. (2007). A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA (New York, N.Y.)*, 13(8):1256–67.

Blount, K. F. and Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature biotechnology*, 24(12):1558–64.

Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N., and Breaker, R. R. (2007). Antibacterial lysine analogs that target lysine riboswitches. *Nature chemical biology*, 3(1):44–9.

Bozdogan, B. and Appelbaum, P. C. (2004). Oxazolidinones: activity, mode of action, and mechanism of resistance. *International journal of antimicrobial agents*, 23(2):113–9.

Brenk, R., Schipani, A., James, D., Krasowski, A., Gilbert, I. H., Frearson, J., and Wyatt, P. G. (2008). Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, 3(3):435–44.

Brenk, R., Vetter, S. W., Boyce, S. E., Goodin, D. B., and Shoichet, B. K. (2006). Probing molecular docking in a charged model binding site. *Journal of molecular biology*, 357(5):1449–70.

Brodersen, D. E., Clemons, W. M., Carter, a. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, 103(7):1143–54.

Buck, J., Fürtig, B., Noeske, J., Wöhnert, J., and Schwalbe, H. (2007). Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40):15699–704.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 5):760–3.

Chao, J. a. and Williamson, J. R. (2004). Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure (London, England : 1993)*, 12(7):1165–76.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. a., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1):12–21.

Clegg, R. M., Murchie, a. I., Zechel, A., Carlberg, C., Diekmann, S., and Lilley, D. M. (1992). Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, 31(20):4846–56.

Cléry, a., Senty-Ségault, V., Leclerc, F., Raué, H. a., and Branlant, C. (2007). Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Molecular and cellular biology*, 27(4):1191–206.

Collins, J. a., Irnov, I., Baker, S., and Winkler, W. C. (2007). Mechanism of mRNA destabilization by the glmS ribozyme. *Genes & development*, 21(24):3356–68.

Corbino, K. a., Barrick, J. E., Lim, J., Welz, R., Tucker, B. J., Puskarz, I., Mandal, M., Rudnick, N. D., and Breaker, R. R. (2005). Evidence for a second

class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome biology*, 6(8):R70.

Croft, M. T., Moulin, M., Webb, M. E., and Smith, A. G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, 104(52):20770–5.

Daldrop, P., Reyes, F. E., Robinson, D. a., Hammond, C. M., Lilley, D. M., Batey, R. T., and Brenk, R. (2011). Novel ligands for a purine riboswitch discovered by RNA-ligand docking. *Chemistry & biology*, 18(3):324–35.

de Beer, S. B. a., Vermeulen, N. P. E., and Oostenbrink, C. (2010). The role of water molecules in computational drug design. *Current topics in medicinal chemistry*, 10(1):55–66.

Deigan, K. E. and Ferre-D'Amare, A. R. (2011). Riboswitches: Discovery of Drugs That Target Bacterial Gene-Regulatory RNAs. *Accounts of chemical research*, (in press).

Delfosse, V., Bouchard, P., Bonneau, E., Dagenais, P., Lemay, J.-F., Lafontaine, D. a., and Legault, P. (2010). Riboswitch structure: an internal residue mimicking the purine ligand. *Nucleic acids research*, 38(6):2057–68.

Demain, A. L. (2009). Antibiotics: natural products essential to human health. *Medicinal research reviews*, 29(6):821–42.

Deng, N.-j. and Cieplak, P. (2010). Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophysical journal*, 98(4):627–36.

Detering, C. and Varani, G. (2004). Validation of automated docking programs for docking and database screening against RNA drug targets. *Journal of medicinal chemistry*, 47(17):4188–201.

Draper, D. E. (2004). A guide to ions and RNA structure. *RNA (New York, N.Y.)*, 10(3):335–43.

Draper, D. E. (2008). RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophysical journal*, 95(12):5489–95.

Dunkle, J. a., Wang, L., Feldman, M. B., Pulk, A., Chen, V. B., Kapral, G. J., Noeske, J., Richardson, J. S., Blanchard, S. C., and Cate, J. H. D. (2011). Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science (New York, N.Y.)*, 332(6032):981–4.

Edwards, A. L. and Batey, R. T. (2009). A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. *Journal of molecular biology*, 385(3):938–48.

Edwards, T. E. and Ferré-D'Amaré, A. R. (2006). Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure (London, England : 1993)*, 14(9):1459–68.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1):2126–32.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4):486–501.

Fedor, M. J. and Williamson, J. R. (2005). The catalytic diversity of RNAs. *Nature reviews. Molecular cell biology*, 6(5):399–412.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology*, 10:226–36.

Franceschi, F. and Duffy, E. M. (2006). Structure-based drug design meets the ribosome. *Biochemical pharmacology*, 71(7):1016–25.

Frankel, a. D. and Young, J. a. (1998). HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry*, 67:1–25.

Fulle, S. and Gohlke, H. (2010). Molecular recognition of RNA: challenges for modelling interactions and plasticity. *Journal of molecular recognition : JMR*, 23(2):220–31.

Garst, A. D., Héroux, A., Rambo, R. P., and Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *The Journal of biological chemistry*, 283(33):22347–51.

Gerber, P. R. and Müller, K. (1995). MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *Journal of computer-aided molecular design*, 9(3):251–68.

Gilbert, S. D., Mediatore, S. J., and Batey, R. T. (2006a). Modified pyrimidines specifically bind the purine riboswitch. *Journal of the American Chemical Society*, 128(44):14214–5.

Gilbert, S. D., Rambo, R. P., Van Tyne, D., and Batey, R. T. (2008). Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature structural & molecular biology*, 15(2):177–82.

Gilbert, S. D., Reyes, F. E., Edwards, A. L., and Batey, R. T. (2009). Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure (London, England : 1993)*, 17(6):857–68.

Gilbert, S. D., Stoddard, C. D., Wise, S. J., and Batey, R. T. (2006b). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of molecular biology*, 359(3):754–68.

Gohlke, C., Murchie, a. I., Lilley, D. M., and Clegg, R. M. (1994). Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24):11660–4.

Goody, T. a., Melcher, S. E., Norman, D. G., and Lilley, D. M. J. (2004). The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA (New York, N.Y.)*, 10(2):254–64.

Gorelik, B. and Goldblum, A. (2008). High quality binding modes in docking ligands to proteins. *Proteins*, 71(3):1373–86.

Gottesman, S. (2002). Stealth regulation: biological circuits with small RNA switches. *Genes & development*, 16(22):2829–42.

Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. (2010). Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of computer-aided molecular design*, 24(6-7):591–604.

Grillo, M. a. and Colombatto, S. (2008). S-adenosylmethionine and its products. *Amino acids*, 34(2):187–93.

Grojean, J. and Downes, B. (2010). Riboswitches as hormone receptors: hypothetical cytokinin-binding riboswitches in Arabidopsis thaliana. *Biology direct*, 5(1):60.

Grundy, F. J. and Henkin, T. M. (1998). The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Molecular microbiology*, 30(4):737–49.

Guay, C., Roggli, E., Nesca, V., Jacovetti, C., and Regazzi, R. (2011). Diabetes mellitus, a microRNA-related disease? *Translational research : the journal of laboratory and clinical medicine*, 157(4):253–64.

Guilbert, C. and James, T. L. (2008). Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of chemical information and modeling*, 48(6):1257–68.

Gusarov, I. and Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Molecular cell*, 3(4):495–504.

Halgren, T. A. (1996). Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6):490–519.

Hamma, T. and Ferré-D'Amaré, A. R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure (London, England : 1993)*, 12(5):893–903.

Hansen, J. L., Ippolito, J. a., Ban, N., Nissen, P., Moore, P. B., and Steitz, T. a. (2002). The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Molecular cell*, 10(1):117–28.

Hartshorn, M. J., Verdonk, M. L., Chessari, G., Brewerton, S. C., Mooij, W. T. M., Mortenson, P. N., and Murray, C. W. (2007). Diverse, high-quality test set for the validation of protein-ligand docking performance. *Journal of medicinal chemistry*, 50(4):726–41.

Hawkey, P. M. (2008). The growing burden of antimicrobial resistance. *The Journal of antimicrobial chemotherapy*, 62 Suppl 1:i1–9.

Hennelly, S. P. and Sanbonmatsu, K. Y. (2010). Tertiary contacts control switching of the SAM-I riboswitch. *Nucleic acids research*, 39(6):2416–31.

Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J. C., and Lafontaine, D. A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature chemical biology*, 7(6):384–92.

Hermann, T. (2000). Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. *Angewandte Chemie (International ed. in English)*, 39(11):1890–1904.

Hermann, T. (2005). Drugs targeting the ribosome. *Current opinion in structural biology*, 15(3):355–66.

Hopkins, A. L., Groom, C. R., and Alex, A. (2004). Ligand efficiency: a useful metric for lead selection. *Drug discovery today*, 9(10):430–1.

Huang, Y., Shen, X. J., Zou, Q., Wang, S. P., Tang, S. M., and Zhang, G. Z. (2011). Biological functions of microRNAs: a review. *Journal of physiology and biochemistry*, 67(1):129–39.

Jenison, R. D., Gill, S. C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science (New York, N.Y.)*, 263(5152):1425–9.

Jovanovic, M. and Hengartner, M. O. (2006). miRNAs and apoptosis: RNAs to die for. *Oncogene*, 25(46):6176–87.

Kalliokoski, T., Salo, H. S., Lahtela-Kakkonen, M., and Poso, A. (2009). The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *Journal of chemical information and modeling*, 49(12):2742–8.

Kang, X., Shafer, R. H., and Kuntz, I. D. (2004). Calculation of ligand-nucleic acid binding free energies with the generalized-born model in DOCK. *Biopolymers*, 73(2):192–204.

Kim, J. N., Blount, K. F., Puskarz, I., Lim, J., Link, K. H., and Breaker, R. R. (2009). Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS chemical biology*, 4(11):915–27.

Kim, R. and Skolnick, J. (2008). Assessment of programs for ligand binding affinity prediction. *Journal of computational chemistry*, 29(8):1316–31.

Klebe, G. (2006). Virtual ligand screening: strategies, perspectives and limitations. *Drug discovery today*, 11(13-14):580–94.

Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. a. (2001). The kink-turn: a new RNA secondary structure motif. *The EMBO journal*, 20(15):4214–21.

Kolb, P., Ferreira, R. S., Irwin, J. J., and Shoichet, B. K. (2009). Docking and chemoinformatic screens for new ligands and targets. *Current opinion in biotechnology*, 20(4):429–36.

Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., Rizzo, R. C., Case, D. A., James, T. L., and Kuntz, I. D. (2009). DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA (New York, N.Y.)*, 15(6):1219–30.

Lee, E. R., Blount, K. F., and Breaker, R. R. (2009). Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA biology*, 6(2):187–94.

Lee, M.-K., Gal, M., Frydman, L., and Varani, G. (2010). Real-time multidimensional NMR follows RNA folding with second resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20):9192–7.

Leipply, D. and Draper, D. E. (2011). Effects of Mg2+ on the free energy landscape for folding a purine riboswitch RNA. *Biochemistry*, 50(14):2790–9.

Lemay, J.-F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Massé, E., and Lafontaine, D. a. (2011). Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS genetics*, 7(1):e1001278.

Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J., and Lafontaine, D. a. (2006). Folding of the adenine riboswitch. *Chemistry & biology*, 13(8):857–68.

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 1):48–57.

Li, L. and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, 443(7109):302–7.

Li, Y., Shen, J., Sun, X., Li, W., Liu, G., and Tang, Y. (2010). Accuracy assessment of protein-based docking programs against RNA targets. *Journal of chemical information and modeling*, 50(6):1134–46.

Liao, C. and Nicklaus, M. C. (2009). Comparison of nine programs predicting pK(a) values of pharmaceutical substances. *Journal of chemical information and modeling*, 49(12):2801–12.

Lilley, D. M. J. (2011). Catalysis by the nucleolytic ribozymes. *Biochemical Society transactions*, 39(2):641–6.

Lim, J., Winkler, W. C., Nakamura, S., Scott, V., and Breaker, R. R. (2006). Molecular-recognition characteristics of SAM-binding riboswitches. *Angewandte Chemie (International ed. in English)*, 45(6):964–8.

Lin, a. H., Murray, R. W., Vidmar, T. J., and Marotti, K. R. (1997). The oxazolidinone eperezolid binds to the 50S ribosomal subunit and competes with binding of chloramphenicol and lincomycin. *Antimicrobial agents and chemotherapy*, 41(10):2127–31.

Lin, J.-C. and Thirumalai, D. (2008). Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *Journal of the American Chemical Society*, 130(43):14080–1.

Lind, K. E., Du, Z., Fujinaga, K., Peterlin, B. M., and James, T. L. (2002). Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target TAR RNA. *Chemistry & biology*, 9(2):185–93.

Lipinski, C. a. (2001). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of pharmacological and toxicological methods*, 44(1):235–49.

Liu, J. and Lilley, D. M. J. (2007). The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *RNA (New York, N.Y.)*, 13(2):200–10.

Loenen, W. a. M. (2006). S-adenosylmethionine: jack of all trades and master of everything? *Biochemical Society transactions*, 34(Pt 2):330–3.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. *Cell*, 139(4):770–9.

Lorber, D. M. and Shoichet, B. K. (1998). Flexible ligand docking using conformational ensembles. *Protein science : a publication of the Protein Society*, 7(4):938–50.

Lu, C., Smith, A. M., Fuchs, R. T., Ding, F., Rajashankar, K., Henkin, T. M., and Ke, A. (2008). Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature structural & molecular biology*, 15(10):1076–83.

Lütticken, R. and Kunstmann, G. (1988). Vancomycin-resistant Streptococcaceae from clinical material. *Zentralblatt für Bakteriologie, Mikrobiologie, und Hygiene. Series A, Medical microbiology, infectious diseases, virology, parasitology*, 267(3):379–82.

Manche, L., Green, S. R., Schmedt, C., and Mathews, M. B. (1992). Interactions between double-stranded RNA regulators and the protein kinase DAI. *Molecular and cellular biology*, 12(11):5238–48.

Manchester, J., Walkup, G., Rivin, O., and You, Z. (2010). Evaluation of pKa estimation methods on 211 druglike compounds. *Journal of chemical information and modeling*, 50(4):565–71.

Mandal, M. and Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology*, 11(1):29–35.

Marco, E. and Gago, F. (2007). Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *ChemMedChem*, 2(10):1388–401.

Matsumura, S., Ikawa, Y., and Inoue, T. (2003). Biochemical characterization of the kink-turn RNA motif. *Nucleic acids research*, 31(19):5544–51.

McClure, W. R. and Cech, C. L. (1978). On the mechanism of rifampicin inhibition of RNA synthesis. *The Journal of biological chemistry*, 253(24):8949–56.

McInnes, C. (2007). Virtual screening strategies in drug discovery. *Current opinion in chemical biology*, 11(5):494–502.

Moitessier, N., Westhof, E., and Hanessian, S. (2006). Docking of aminoglycosides to hydrated and flexible RNA. *Journal of medicinal chemistry*, 49(3):1023–33.

Montange, R. K. and Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097):1172–5.

Montange, R. K., Mondragón, E., van Tyne, D., Garst, A. D., Ceres, P., and Batey, R. T. (2010). Discrimination between closely related cellular metabolites by the SAM-I riboswitch. *Journal of molecular biology*, 396(3):761–72.

Morley, S. D. and Afshar, M. (2004). Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *Journal of computer-aided molecular design*, 18(3):189–208.

Mpamhanga, C. P., Spinks, D., Tulloch, L. B., Shanks, E. J., Robinson, D. a., Collie, I. T., Fairlamb, A. H., Wyatt, P. G., Frearson, J. a., Hunter, W. N., Gilbert, I. H., and Brenk, R. (2009). One scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. *Journal of medicinal chemistry*, 52(14):4454–65.

Mulhbacher, J. and Lafontaine, D. a. (2007). Ligand recognition determinants of guanine riboswitches. *Nucleic acids research*, 35(16):5568–80.

Mulhbacher, J., St-Pierre, P., and Lafontaine, D. a. (2010). Therapeutic applications of ribozymes and riboswitches. *Current opinion in pharmacology*, 10(5):551–6.

Muniyappa, M. K., Dowling, P., Henry, M., Meleady, P., Doolan, P., Gammell, P., Clynes, M., and Barron, N. (2009). MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines. *European journal of cancer (Oxford, England : 1990)*, 45(17):3104–18.

Mysinger, M. M. and Shoichet, B. K. (2010). Rapid context-dependent ligand desolvation in molecular docking. *Journal of chemical information and modeling*, 50(9):1561–73.

Nahvi, A., Barrick, J. E., and Breaker, R. R. (2004). Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic acids research*, 32(1):143–50.

Neupane, K., Yu, H., Foster, D. a. N., Wang, F., and Woodside, M. T. (2011). Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic acids research*, 780:1–11.

Nicholls, A. and Honig, B. (1991). A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 12(4):435–445.

Nudler, E. and Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. *Trends in biochemical sciences*, 29(1):11–7.

Ott, E., Stolz, J., Lehmann, M., and Mack, M. (2009). The RFN riboswitch of Bacillus subtilis is a target for the antibiotic roseoflavin produced by Streptomyces davawensis. *RNA biology*, 6(3):276–80.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276(Macromolecular Crystallography, Part A):307–326.

Ouellet, J., Melcher, S., Iqbal, A., Ding, Y., and Lilley, D. M. J. (2010). Structure of the three-way helical junction of the hepatitis C virus IRES element. *RNA (New York, N.Y.)*, 16(8):1597–609.

Park, S.-J., Jung, Y. H., Kim, Y.-G., and Park, H.-J. (2008). Identification of novel ligands for the RNA pseudoknot that regulate -1 ribosomal frameshifting. *Bioorganic & medicinal chemistry*, 16(8):4676–84.

Pfeffer, P. and Gohlke, H. (2007). DrugScoreRNA–knowledge-based scoring function to predict RNA-ligand interactions. *Journal of chemical information and modeling*, 47(5):1868–76.

Pinto, I. G., Guilbert, C., Ulyanov, N. B., Stearns, J., and James, T. L. (2008). Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. *Journal of medicinal chemistry*, 51(22):7205–15.

Poiata, E., Meyer, M. M., Ames, T. D., and Breaker, R. R. (2009). A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA (New York, N.Y.)*, 15(11):2046–56.

Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chemistry & biology*, 17(5):421–33.

Rieder, R., Lang, K., Graber, D., and Micura, R. (2007). Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem : a European journal of chemical biology*, 8(8):896–902.

Ripphausen, P., Nisius, B., Peltason, L., and Bajorath, J. (2010). Quo vadis, virtual screening? A comprehensive survey of prospective applications. *Journal of medicinal chemistry*, 53(24):8461–7.

Rodionov, D. a., Vitreschak, A. G., Mironov, A. a., and Gelfand, M. S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *The Journal of biological chemistry*, 277(50):48949–59.

Rodionov, D. a., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2003). Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic acids research*, 31(23):6748–57.

Rodionov, D. a., Vitreschak, A. G., Mironov, A. a., and Gelfand, M. S. (2004). Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic acids research*, 32(11):3340–53.

Rodnina, M. V., Beringer, M., and Wintermeyer, W. (2007). How ribosomes make peptide bonds. *Trends in biochemical sciences*, 32(1):20–6.

Rosenfeld, R. J., Goodsell, D. S., Musah, R. a., Morris, G. M., Goodin, D. B., and Olson, A. J. (2003). Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *Journal of computer-aided molecular design*, 17(8):525–36.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102(5):615–23.

Schneider, G. (2010). Virtual screening: an endless staircase? *Nature reviews. Drug discovery*, 9(4):273–6.

Schroeder, K. T. and Lilley, D. M. J. (2009). Ion-induced folding of a kink turn that departs from the conventional sequence. *Nucleic acids research*, 37(21):7281–9.

Schroeder, K. T., McPhee, S. a., Ouellet, J., and Lilley, D. M. J. (2010). A structural database for k-turn motifs in RNA. *RNA (New York, N.Y.)*, 16(8):1463–8.

Schüttelkopf, A. W. and van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 8):1355–63.

Schwalbe, H., Buck, J., Fürtig, B., Noeske, J., and Wöhnert, J. (2007). Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angewandte Chemie (International ed. in English)*, 46(8):1212–9.

Sen, G. L. and Blau, H. M. (2006). A brief history of RNAi: the silence of the genes. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 20(9):1293–9.

Serganov, A. (2010). Determination of riboswitch structures: light at the end of the tunnel? *RNA biology*, 7(1):98–103.

Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., and Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & biology*, 11(12):1729–41.

Shoichet, B. K., Dale, B. L., and Kuntz, I. D. (1992). Molecular docking using shape descriptors. *Journal of Computational Chemistry*, 103(3):227–397.

Sledz, C. a., Holko, M., de Veer, M. J., Silverman, R. H., and Williams, B. R. G. (2003). Activation of the interferon system by short-interfering RNAs. *Nature cell biology*, 5(9):834–9.

Smith, K. D., Lipchock, S. V., Ames, T. D., Wang, J., Breaker, R. R., and Strobel, S. a. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. *Nature structural & molecular biology*, 16(12):1218–23.

Smith, K. D., Shanahan, C. a., Moore, E. L., Simon, A. C., and Strobel, S. a. (2011). Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19):7757–62.

Stjernschantz, E. and Oostenbrink, C. (2010). Improved ligand-protein binding affinity predictions using multiple binding modes. *Biophysical journal*, 98(11):2682–91.

Stoddard, C. D., Montange, R. K., Hennelly, S. P., Rambo, R. P., Sanbonmatsu, K. Y., and Batey, R. T. (2010). Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure (London, England : 1993)*, 18(7):787–97.

Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003a). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA (New York, N.Y.)*, 9(6):644–7.

Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H., and Breaker, R. R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science (New York, N.Y.)*, 321(5887):411–3.

Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S., and Breaker, R. R. (2003b). An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & development*, 17(21):2688–97.

ten Brink, T. and Exner, T. E. (2009). Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *Journal of chemical information and modeling*, 49(6):1535–46.

Thomas, J. R. and Hergenrother, P. J. (2008). Targeting RNA with small molecules. *Chemical reviews*, 108(4):1171–224.

Thore, S., Frick, C., and Ban, N. (2008). Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch. *Journal of the American Chemical Society*, 130(26):8116–7.

Tomsic, J., McDaniel, B. a., Grundy, F. J., and Henkin, T. M. (2008). Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. *Journal of bacteriology*, 190(3):823–33.

Tucker, B. J. and Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Current opinion in structural biology*, 15(3):342–8.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968):505–10.

Turnbull, W. B. (2005). Divided We Fall ? Studying low affinity fragments of ligands by ITC Written by :. *Microcal Application note*, 1:1–10.

Turner, B. and Lilley, D. M. J. (2008). The importance of G.A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. *Journal of molecular biology*, 381(2):431–42.

Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G., and Lilley, D. M. J. (2005). Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA (New York, N.Y.)*, 11(8):1192–200.

Vagin, A. and Teplyakov, A. (1997). MOLREP : an Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, 30(6):1022–1025.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1):2184–95.

Verdonk, M. L., Berdini, V., Hartshorn, M. J., Mooij, W. T. M., Murray, C. W., Taylor, R. D., and Watson, P. (2004). Virtual screening using protein-ligand docking: avoiding artificial enrichment. *Journal of chemical information and computer sciences*, 44(3):793–806.

Vicens, Q., Mondragón, E., and Batey, R. T. (2011). Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic acids research*, 39(19):8586–8598.

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends in genetics : TIG*, 20(1):44–50.

Waksman, S. A. (1947). What is an antibiotic or an antibiotic substance? *Mycologia*, 39(5):565–9.

Walter, F., Pütz, J., Giegé, R., and Westhof, E. (2002). Binding of tobramycin leads to conformational changes in yeast tRNA(Asp) and inhibition of aminoacylation. *The EMBO journal*, 21(4):760–8.

Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049.

Wang, J. X. and Breaker, R. R. (2008). Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 86(2):157–68.

Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W., and Shoichet, B. K. (2002). A model binding site for testing scoring functions in molecular docking. *Journal of molecular biology*, 322(2):339–55.

Weinberg, Z., Regulski, E. E., Hammond, M. C., Barrick, J. E., Yao, Z., Ruzzo, W. L., and Breaker, R. R. (2008). The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA (New York, N.Y.)*, 14(5):822–8.

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf Comput. Sci.*, 28(1):31–36.

Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R. M., Petry, S., Kelley, A. C., and Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. *Science (New York, N.Y.)*, 322(5903):953–6.

Wickiser, J. K., Cheah, M. T., Breaker, R. R., and Crothers, D. M. (2005). The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, 44(40):13404–14.

Wilson, D. S. and Szostak, J. W. (1999). In vitro selection of functional nucleic acids. *Annual review of biochemistry*, 68:611–47.

Wilson, T. J., Zhao, Z. Y., Maxwell, K., Kontogiannis, L., and Lilley, D. M. (2001). Importance of specific nucleotides in the folding of the natural form of the hairpin ribozyme. *Biochemistry*, 40(7):2291–302.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, a. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407(6802):327–39.

Winkler, W. C., Cohen-Chalamish, S., and Breaker, R. R. (2002). An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25):15908–13.

Winkler, W. C., Nahvi, A., Roth, A., Collins, J. a., and Breaker, R. R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980):281–6.

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature structural biology*, 10(9):701–7.

Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical biochemistry*, 179(1):131–7.

Woźniak, A. K., Nottrott, S., Kühn-Hölsken, E., Schröder, G. F., Grubmüller, H., Lührmann, R., Seidel, C. a. M., and Oesterhelt, F. (2005). Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA (New York, N.Y.)*, 11(10):1545–54.

Yan, K., Hunt, E., Berge, J., May, E., Copeland, R. A., and Gontarek, R. R. (2005). Fluorescence polarization method to characterize macrolide-ribosome interactions. *Antimicrobial agents and chemotherapy*, 49(8):3367–72.

Yarnell, W. S. and Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science (New York, N.Y.)*, 284(5414):611–5.

Yoshizawa, S., Fourmy, D., and Puglisi, J. D. (1999). Recognition of the codon-anticodon helix by ribosomal RNA. *Science (New York, N.Y.)*, 285(5434):1722–5.

Yuriev, E., Agostino, M., and Ramsland, P. a. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of molecular recognition : JMR*, 24(2):149–64.

Zhang, Y. L. and Zhang, Z. Y. (1998). Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. *Analytical biochemistry*, 261(2):139–48.